

GenCore version 5.1.6  
Copyright (C) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 19:57:16 ; Search time 79.2548 Seconds  
(without alignments)  
3347.654 Million cell updates/sec

Title: US-10-658-782-2

Perfect score: 3619

Sequence: 1 MAPITAYAQTRGLGLCIIT.....PAIPDREVLRYFDEMEEC 686

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3619	100.0	686	5	Aau76377 Hepatitis
2	3619	100.0	686	5	Aae18689 HCV-1 NS3
3	3619	100.0	686	6	Abg72261 HCV-1 NS3
4	3619	100.0	686	7	Adc06767 HCV mutan
5	3619	100.0	686	8	Adl66805 HCV NS3/4
6	3602	99.5	2261	1	Aap90164 Peptide e
7	3602	99.5	2436	1	Aap92050 HCV prote
8	3602	99.5	2436	1	Aap90288 Peptide e
9	3602	99.5	2772	3	Aab18540 Protein e
10	3602	99.5	2772	8	Adn35976 HCV cDNA
11	3602	99.5	2955	2	Aay14975 Amino aci
12	3602	99.5	2955	3	Aab18541 Polypept
13	3602	99.5	2955	8	Adn35978 HCV cDNA
14	3602	99.5	3011	2	Aar90931 Hepatitis
15	3602	99.5	3011	2	Aaw34480 HCV polyp
16	3602	99.5	3011	2	Aaw40038 HCV polyp
17	3602	99.5	3011	5	Aae22049 Hepatitis
18	3602	99.5	3011	8	Adl23107 Hepatitis
19	3602	99.5	3011	8	Adr29357 Hepatitis
20	3600	99.5	728	5	Aae18688 NS3/4a mu
21	3600	99.5	728	7	Adc06766 HCV mutan
22	3599	99.4	2301	1	Aap92047 HCV prote
23	3595	99.3	2772	2	Aar08123 Hepatitis
24	3594	99.3	686	4	Aab62633 HCV NS3/4
25	3593	99.3	3011	2	Aar21519 Compiled

26	3590	99.2	2435	2	AAR25135	Aar25135 HCV polyp
27	3589	99.2	3011	2	AAR31621	Aar31621 Hepatitis
28	3587	99.1	3011	5	AAU84597	Aau84597 HCV polyp
29	3586	99.1	2816	2	AAR34009	Aar34009 HCV-1 pol
30	3583	99.0	1786	1	AAP90158	Aap90158 Protein 8
31	3583	99.0	2436	2	AAR28582	Aar28582 HCV amino
32	3583	99.0	2894	2	AAR70230	Aar70230 Compoite
33	3580	98.9	2894	2	AAR24440	Aar24440 Compoite
34	3579	98.9	1766	1	AAP92041	Aap92041 Hepatitis
35	3565	98.5	686	5	AAE21837	Aae21837 Hepatitis
36	3565	98.5	686	5	AAE19900	Aae19900 Hepatitis
37	3565	98.5	686	7	ABW00351	Abw00351 Hepatitis
38	3565	98.5	686	8	ADG47659	Adg47659 HCV NS3/4
39	3565	98.5	686	8	ADG47693	Adg47693 HCV NS3/4
40	3563	98.5	1986	8	ADR38451	Adr38451 Hepatitis
41	3561	98.4	686	5	AAE21838	Aae21838 Hepatitis
42	3561	98.4	686	5	AAE19907	Aae19907 Hepatitis
43	3561	98.4	686	7	ABW00358	Abw00358 Hepatitis
44	3561	98.4	686	8	ADG47660	Adg47660 HCV NS3/4
45	3561	98.4	3011	2	AAR40120	Aar40120 HCV genom

#### ALIGNMENTS

RESULT 1  
AAU76377  
ID AAU76377 standard; protein; 686 AA.  
XX  
AC AAU76377;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Hepatitis C virus NS3/4a conformational epitope protein sequence.  
XX  
KW Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
KW immunoassay solid support; multiple epitope fusion antigen; MEPA;  
KW non-structural protein; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 403 /note= "Wild-type Thr substituted by Pro"  
FT Misc-difference 404 /note= "Wild-type Ser substituted by Ile"  
XX  
PN WO200196870-A2.  
XX  
PD 20-DEC-2001.  
XX  
PF 14-JUN-2001; 2001WO-US019156.  
XX  
PR 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX  
WPI; 2002-090228/12.  
DR N-PSDB; ABK15344.  
XX  
PT Immunoassay solid support, useful for detecting hepatitis C virus  
PT infection in biological sample, comprises HCV NS3/4a conformational  
PT epitope and multiple epitope fusion antigen bound to the support.  
XX  
PS Claim 5; Fig 3; 92pp; English.  
XX  
CC The present invention relates to a new immunoassay solid support





Qy 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 540  
 Db 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 540  
 Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600  
 Db 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600  
 Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCISGCVIVGR 660  
 Db 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCISGCVIVGR 660  
 Qy 661 VVLSGKPAIIPDREVLVREFDEMEEC 686  
 Db 661 VVLSGKPAIIPDREVLVREFDEMEEC 686

RESULT 4  
 ADC06767  
 ID ADC06767 standard; protein; 686 AA.  
 AC  
 XX  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE HCV mutant conformational NS3/4a epitope protein T403P/S404I.  
 XX  
 KW immunoassay solid support; HCV; NS3/4a; non-structural;  
 KW non-A, non-B hepatitis; NANB; conformational epitope; mutant; mutein.  
 XX  
 OS Synthetic.  
 OS Hepatitis C virus.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 303  
 FT /note= "Wild-type Thr replaced by Pro"  
 FT Misc-difference 304  
 FT /note= "Wild-type Ser replaced by Ile"  
 XX  
 PN US2002192639-A1.  
 XX  
 PD 19-DEC-2002.  
 XX  
 PF 14-JUN-2001; 2001US-00881239.  
 XX  
 PR 15-JUN-2000; 2000US-0212082P.  
 PR 02-APR-2001; 2001US-0280811P.  
 PR 02-APR-2001; 2001US-0280867P.  
 XX  
 PA (CHIE/) CHIEN D Y.  
 PA (ARCA/) ARCANGEL P.  
 PA (TAND/) TANDESKE L.  
 PA (GEOR/) GEORGE-NASCIMENTO C.  
 PA (COIT/) COIT D.  
 PA (MEDI/) MEDINA-SELBY A.  
 XX  
 PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
 PI Medina-Selby A;  
 XX  
 DR WPI; 2003-644609/61.  
 XX N-PSDB; ADC06768.  
 XX  
 PT Immunoassay solid support for detecting hepatitis C virus infection in  
 PT biological samples, comprises a hepatitis C virus anti-core antibody and  
 PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
 PT antibody.  
 XX  
 PS Claim 6; Fig 4; 40pp; English.  
 XX  
 CC The invention relates to a novel immunoassay solid support comprising at  
 CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
 CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.

CC The system of the invention may be useful for detecting HCV infection in  
 CC a biological sample and for treating or detecting non-A, non-B hepatitis  
 CC (NANB hepatitis). The current sequence is that of the HCV mutant  
 CC conformational NS3/4a epitope protein of the invention which contains  
 CC T403P/S404I mutations.

SQ Sequence 686 AA;

Query Match 100.0%; Score 3619; DB 7; Length 686;

Best Local Similarity 100.0%; Pred. No. 2.5e-306;

Matches 686; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60

Db 1 MAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60

Qy 61 GTRTIASPKGPVIQMYTNVDQDLVGNPAPQGSRLTPTCTCGSSDLVLTVRHADVIPVRR 120

Db 61 GTRTIASPKGPVIQMYTNVDQDLVGNPAPQGSRLTPTCTCGSSDLVLTVRHADVIPVRR 120

Qy 121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180

Db 121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180

Qy 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAVAAQYKVLNPSVAATLGF 240

Db 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAVAAQYKVLNPSVAATLGF 240

Qy 241 AYMSKAHGIDPNIRTGVRTITTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300

Db 241 AYMSKAHGIDPNIRTGVRTITTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGVSUTVPHNIEEVALSTTGEIPYGAIPLEVI 360

Db 301 ILGIGTVLDOAETAGARLVVLTATPPGVSUTVPHNIEEVALSTTGEIPYGAIPLEVI 360

Qy 361 KGRHLIFCHSKKCKDELAALVALGINAVAYYRGDVSVIPIGDVVVVATDALMTGYT 420

Db 361 KGRHLIFCHSKKCKDELAALVALGINAVAYYRGDVSVIPIGDVVVVATDALMTGYT 420

Qy 421 GDFDSVIDCNTCTVTQVDFSLDPTFTIETITLPODAVSRRTQRRGTRGKPGIYRVFAPG 480

Db 421 GDFDSVIDCNTCTVTQVDFSLDPTFTIETITLPODAVSRRTQRRGTRGKPGIYRVFAPG 480

Qy 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 540

Db 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 540

Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600

Db 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600

Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCISGCVIVGR 660

Db 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCISGCVIVGR 660

Qy 661 VVLSGKPAIIPDREVLVREFDEMEEC 686

Db 661 VVLSGKPAIIPDREVLVREFDEMEEC 686

RESULT 5

ADL6805

ID ADL6805 standard; protein; 686 AA.

XX ADL6805;

XX ADL6805;

DT 03-JUN-2004 (first entry)

XX HCV NS3/4a conformational epitope.

XX HCV; NS3/4a conformational epitope; HCV antigen; HCV polyprotein;

KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection.



```
XX OS Hepatitis C virus.
XX PN WO2004021871-A2.
XX XX
XX PD 18-MAR-2004.
XX PF
XX PR 08-SEP-2003; 2003WO-US028071.
XX PR 09-SEP-2002; 2002US-0409515P.
XX PA (CHIR ) CHIRON CORP.
XX PI Arcangel P, Chien D;
XX PR WPI; 2004-248333/23.
XX DR N-PSDB; ADL66804.
XX XX
XX PT Detecting hepatitis C virus (HCV) infection in a biological sample by
XX PT detecting complexes formed between the HCV antibody and the antigens from
XX PT the first region of the HCV polyprotein and the multiple epitope fusion
XX PT antigen (MEFA).
XX XX
XX PS Claim 5; SEQ ID NO 2; 93pp; English.
XX XX
XX CC The invention relates to a method of detecting hepatitis C virus (HCV)
XX CC infection in a biological sample. The method comprises providing an
XX CC immunocassay solid support comprising HCV antigens bound to it, where the
XX CC HCV antigens comprise one or more isolated antigens form a first region
XX CC of the HCV polyprotein, combining a biological sample with the solid
XX CC support under conditions that allow HCV antibodies, when present in the
XX CC biological sample, to bind to the one or more HCV antigens, adding to the
XX CC solid support a detectably labelled HCV multiple epitope fusion antigen
XX CC (MEFA), where the labelled MEFA comprises at least one epitope from the
XX CC same region of the HCV polyprotein as the one or more isolated antigens,
XX CC where the MEFA binds to the bound HCV antibody, and detecting complexes
XX CC formed between the HCV antibody and the one or more antigens from the
XX CC first region of the HCV polyprotein and the MEFA, if any, as an
XX CC indication of HCV infection in the biological sample. The method is
XX CC useful for detecting hepatitis C virus (HCV) infection in a biological
XX CC sample. This sequence represents the NS3/4a conformational epitope used
XX CC in the scope of the invention.
XX SQ Sequence 686 AA;

Query Match 100.0%; Score 3619; DB 8; Length 686;
Best Local Similarity 100.0%; Pred. No. 2.5e-306; Mismatches 0; Gaps 0;
Matches 686; Conservative 0; Indels 0;

Qy 1 MAPITAYAAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60
Db 1 MAPITAYAAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60

Qy 61 GTRTIASPKGVIOYNTNVDDOLGVGPAPQSGRSRLTPTCTGSSDLYLVTRHADVIPVRRR 120
Db 61 GTRTIASPKGVIOYNTNVDDOLGVGPAPQSGRSRLTPTCTGSSDLYLVTRHADVIPVRRR 120

Qy 121 GDSRGLSPRPISVLKSGSGPLLCFAGHAGVIFRAACVTRGKAVKAVDFIPVENLETTM 180
Db 121 GDSRGLSPRPISVLKSGSGPLLCFAGHAGVIFRAACVTRGKAVKAVDFIPVENLETTM 180

Qy 181 RSPVFTDNSSPPVPSQFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240
Db 181 RSPVFTDNSSPPVPSQFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240

Qy 241 AYMSKAHGIDNIRTVGRTITTTGSPITVSTYTKFLADGGCGGGAYDIIICDECHSTDATS 300
Db 241 AYMSKAHGIDNIRTVGRTITTTGSPITVSTYTKFLADGGCGGGAYDIIICDECHSTDATS 300

Qy 301 ILGIGTVLDQAEATAGARLVVLATATPPGSGVTPVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db 301 ILGIGTVLDQAEATAGARLVVLATATPPGSGVTPVPHNIEEVALSTTGEIPFYGKAIPLEVI 360

361 KGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPPIGDVVVVVATDALMTGYT 420
361 KGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPPIGDVVVVVATDALMTGYT 420
421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRRTORRGRTGRGPKGIYRFVAPG 480
421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRRTORRGRTGRGPKGIYRFVAPG 480
481 ERPSGMFSSVLCCEYDAGCAWYELTPTAETTVRLRAYMNTTGLPVCQDHLFEFWEVFTGL 540
481 ERPSGMFSSVLCCEYDAGCAWYELTPTAETTVRLRAYMNTTGLPVCQDHLFEFWEVFTGL 540
541 THIDAHLFSLQSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKPETHLHGPTPLL 600
541 THIDAHLFSLQSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKPETHLHGPTPLL 600
601 YRLGAVONEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
601 YRLGAVONEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
661 VLSGKPAIIPDREVLRYREFDEMEEC 686
661 VLSGKPAIIPDREVLRYREFDEMEEC 686

RESULT 6
AAP90164
ID AAP90164 standard; protein; 2261 AA.
XX AC AAP90164;
XX DT 25-MAR-2003 (revised)
XX DT 01-NOV-1989 (first entry)
XX DE Peptide encoded by composite hepatitis C virus cDNA.
XX XX Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.
XX OS Pan troglodytes.
XX PN GB2212511-A.
XX XX
XX PD 26-JUL-1989.
XX XX
XX PF 18-NOV-1988; 88GB-00027024.
XX PR 18-NOV-1987; 87US-00122714.
XX PR 30-DEC-1987; 87US-00139886.
XX PR 26-FEB-1988; 88US-00161072.
XX PR 26-OCT-1988; 88US-00263584.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo QL, Kuo G;
XX XX WPI; 1989-215054/30.
XX DR N-PSDB; AAN90331.
XX XX Hepatitis C virus gene - used for prodn. of polynucleotide probes
XX PT polypeptide(s) and antibodies for diagnosis, prevention and treatment of
XX PT infection.
XX PS Disclosure; Fig 32; 30pp; English.
XX CC The sequence is the peptide encoded by the composite hepatitis C virus
XX CC (HCV) cDNA of AAN90331. The polypeptides are used to diagnose HCV-induced
XX CC NANBH, to raise antibodies for immunoassay or treatment, or to produce
XX CC vaccines. (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 2261 AA;

Query Match 99.5%; Score 3602; DB 1; Length 2261;
Best Local Similarity 99.6%; Pred. No. 4.4e-304;
```

Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db	401 LAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 460
Qy	61 GTRTIASPKGPVQMTYNDQDLVGHAPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPIVRRR 120
Db	461 GTRTIASPKGPVQMTYNDQDLVGHAPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPIVRRR 520
Qy	121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db	521 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 580
Qy	181 RSPVFTDSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNPSVAATLGF 240
Db	581 RSPVFTDSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNPSVAATLGF 640
Qy	241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300
Db	641 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 700
Qy	301 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db	701 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 760
Qy	361 KGRHLIFCHSKKCCDELAALVALGINAVAYYRGDLDSVIPPIGDVVVATDALMTGYT 420
Db	761 KGRHLIFCHSKKCCDELAALVALGINAVAYYRGDLDSVIPPIGDVVVATDALMTGYT 820
Qy	421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRTGRGKPGIYRFVAPG 480
Db	821 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRTGRGKPGIYRFVAPG 880
Qy	481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 540
Db	881 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 940
Qy	541 THIDAHFLSQTOSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKPETHLHGPTELL 600
Db	941 THIDAHFLSQTOSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKPETHLHGPTELL 1000
Qy	601 YRLGAVQNEITLTHPVTKYMTCMSADLEVTSTWLVGGVLAALAAAYCISLSTGCWTVGR 660
Db	1001 YRLGAVQNEITLTHPVTKYMTCMSADLEVTSTWLVGGVLAALAAAYCISLSTGCWTVGR 1060
Qy	661 VVLGSKPAIIPDREVLRYFDEMEEC 686
Db	1061 VVLGSKPAIIPDREVLRYFDEMEEC 1086
RESULT 7	
AAP92050	
ID	AAP92050 standard; protein; 2436 AA.
XX	
AC	AAP92050;
XX	
DT	09-SEP-2004 (revised)
DT	25-MAR-2003 (revised)
DT	02-MAR-1990 (first entry)
XX	
DE	HCV protein of the cDNA inserts in clones K9-1 through 15e.
XX	
KW	Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).
XX	
OS	Hepatitis C virus.
OS	Unidentified.
XX	
PN	EP318216-A.
XX	
PD	31-MAY-1989.
XX	
PF	18-NOV-1988; 88EP-00310922.

XX	18-NOV-1987; 87US-00122714.
PR	30-DEC-1987; 87US-00139886.
PR	26-FEB-1988; 88US-00161072.
PR	06-MAY-1988; 88US-00181263.
PR	26-OCT-1988; 88US-00263584.
PR	14-NOV-1988; 88US-00271450.
XX	
PA	(CHIR ) CHIRON CORP.
PA	(CHIR ) CHIRON CORP.
XX	
PI	Houghton M, Choo QL, Kuo G;
XX	
DR	WPI; 1989-159274/22.
DR	N-PSDB; AAN92106.
XX	
PT	Purified hepatitis C virus - and associated nucleic acids and
PT	polypeptide(s).
XX	
PS	Claim 13; Fig 47-1-47-8; 139pp; English.
XX	
CC	It is the sequence encoded in the open reading frame of hepatitis C virus
CC	(HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
CC	be used in immunoassay reagents and vaccines and to generate antibodies
CC	useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC	non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC	on 25-MAR-2003 to correct PI field.)
CC	
CC	Revised record issued on 09-SEP-2004 : Correction to DE line
XX	
SQ	Sequence 2436 AA;
Query Match 99.5%; Score 3602; DB 1; Length 2436;	
Best Local Similarity 99.6%; Pred. No. 5e-304;	
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db	576 LAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635
Qy	61 GTRTIASPKGPVQMTYNDQDLVGHAPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPIVRRR 120
Db	636 GTRTIASPKGPVQMTYNDQDLVGHAPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPIVRRR 695
Qy	121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db	696 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755
Qy	181 RSPVFTDSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNPSVAATLGF 240
Db	756 RSPVFTDSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNPSVAATLGF 815
Qy	241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300
Db	816 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 875
Qy	301 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db	876 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 935
Qy	361 KGRHLIFCHSKKCCDELAALVALGINAVAYYRGDLDSVIPPIGDVVVATDALMTGYT 420
Db	936 KGRHLIFCHSKKCCDELAALVALGINAVAYYRGDLDSVIPPIGDVVVATDALMTGYT 995
Qy	421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRTGRGKPGIYRFVAPG 480
Db	996 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRTGRGKPGIYRFVAPG 1055
Qy	481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 540
Db	1056 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 1115
Qy	541 THIDAHFLSQTOSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKPETHLHGPTELL 600

```

Db      1116 THIDAHFLSQTQSGENLPYLVAQVATVCARAQAPPSWDMKCLIRLKPTHLHGPTPLL 1175
Qy      601 YRLGAVQNEITLTHPVTKYIMTCSADLEVTWLVGGVLAALAAAYCLSTGCWIVGR 660
Db      1176 YRLGAVQNEITLTHPVTKYIMTCSADLEVTWLVGGVLAALAAAYCLSTGCWIVGR 1235
Qy      661 VVLSGKPAIIPDREVLRYEFDEMEEC 686
Db      1236 VVLSGKPAIIPDREVLRYEFDEMEEC 1261

RESULT 8
AAP90288
ID AAP90288 standard; protein; 2436 AA.
XX AC AAP90288;
XX DT 25-MAR-2003 (revised)
XX DT 19-JUL-2001 (revised)
XX DT 01-NOV-1989 (first entry)
XX XX
XX Peptide encoded by composite hepatitis C cDNA.
XX Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
XX Pan troglodytes.
XX OS
XX GB2212511-A.
XX PD
XX PF 18-NOV-1988; 88GB-00027024.
XX PR 18-NOV-1987; 87US-00122714.
XX PR 30-DEC-1987; 87US-00139886.
XX PR 26-FEB-1988; 88US-00161072.
XX PR 26-OCT-1988; 88US-00263584.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo QI, Kuo G;
XX WPI; 1989-215054/30.
XX DR N-PSDB; AAN90336.
XX XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes
XX polypeptide(s) and antibodies for diagnosis, prevention and treatment of
XX infection.
XX PS Disclosure; Fig 47-1 to 47-8; 30pp; English.
XX CC The sequence is the peptide encoded by the composite hepatitis C virus
XX (HCV) cDNA of AAN90336. The polypeptides are used to diagnose HCV-induced
XX NANBH, to raise antibodies for immunoassay or treatment, or to produce
XX vaccines. (N.B. This record was resubmitted to correct errors in the
XX sequence.) (Updated on 25-MAR-2003 to correct PR field.)
XX XX
XX SQ Sequence 2436 AA;

Query Match 99.5%; Score 3602; DB 1; Length 2436;
Best Local Similarity 99.6%; Pred. No. 5e-304;
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MAPITAAQQTGRLGCGIITSLTGRDKNQVEGEQIVSTAQTFLATCINGCVCTVYHGA 60
Db      576 LAPITAAQQTGRLGCGIITSLTGRDKNQVEGEQIVSTAQTFLATCINGCVCTVYHGA 635
Qy      61 GTRTIASPKGPVQMYTNVDDQLVGPAPQGSRLTPTCTGSSDLVLVTRHADVIPVRRR 120
Db      636 GTRTIASPKGPVQMYTNVDDQLVGPAPQGSRLTPTCTGSSDLVLVTRHADVIPVRRR 695
Qy      121 GDSRGSLLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180

```

```

Db      696 GDSRGSLLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755
Qy      181 RSPVFTDSSPPVPOSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLNPSVAATLGFG 240
Db      756 RSPVFTDSSPPVPOSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLNPSVAATLGFG 815
Qy      241 AYMSKAHGDIPNIRTVRTITTCSPITYSTYTGKFLADGGCGSGAYDIIICDECHSTDATS 300
Db      816 AYMSKAHGDIPNIRTVRTITTCSPITYSTYTGKFLADGGCGSGAYDIIICDECHSTDATS 875
Qy      301 ILGIGTVLDOAETAGARLVVLAATATPPGSVTVPHNIEEVALSTTGEIPYGAIPLEVI 360
Db      876 ILGIGTVLDOAETAGARLVVLAATATPPGSVTVPHNIEEVALSTTGEIPYGAIPLEVI 935
Qy      361 KGGRHILFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 420
Db      936 KGGRHILFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 995
Qy      421 GDFDSVIDCNTCTVTQTVDFSLDPTFTIETITLQDAVSRTORRGTGRGKPGIYRFVAPG 480
Db      996 GDFDSVIDCNTCTVTQTVDFSLDPTFTIETITLQDAVSRTORRGTGRGKPGIYRFVAPG 1055
Qy      481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFEGVFTGL 540
Db      1056 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFEGVFTGL 1115
Qy      541 THIDAHFLSQTQSGENLPYLVAQVATVCARAQAPPSWDMKCLIRLKPTHLHGPTPLL 600
Db      1116 THIDAHFLSQTQSGENLPYLVAQVATVCARAQAPPSWDMKCLIRLKPTHLHGPTPLL 1175
Qy      601 YRLGAVQNEITLTHPVTKYIMTCSADLEVTWLVGGVLAALAAAYCLSTGCWIVGR 660
Db      1176 YRLGAVQNEITLTHPVTKYIMTCSADLEVTWLVGGVLAALAAAYCLSTGCWIVGR 1235
Qy      661 VVLSGKPAIIPDREVLRYEFDEMEEC 686
Db      1236 VVLSGKPAIIPDREVLRYEFDEMEEC 1261

RESULT 9
AAB18540
ID AAB18540 standard; protein; 2772 AA.
XX AC AAB18540;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX KW viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-00109602.
XX PR 17-MAR-1989; 89US-00325338.
XX PR 20-APR-1989; 89US-00341334.
XX PR 18-MAY-1989; 89US-00355002.
XX PR 16-MAR-1990; 90EP-00302866.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX WPI; 2000-566891/53.
XX DR N-PSDB; AAA75296.
XX XX

```

PT Novel composition comprising a hepatitis C virus antisense polynucleotide  
PT which is complementary to or corresponds to a sense strand of the virus  
PT genome, and selectively hybridizes to it.

PS Example; Fig 16; 75pp; English.

XX The specification describes a pharmaceutical composition which comprises  
CC a hepatitis C virus (HCV) antisense polynucleotide. The HCV is  
CC characterized by a positive stranded RNA genome which has 40% homology at  
CC the polypeptide level to a HCV polypeptide. The antisense polynucleotide  
CC binds to cellular polynucleotides which enhance and/or are required for  
CC viral infectivity, replicative ability or chronicity. The antisense  
CC polynucleotides may also be designed to bind with high specificity, to be  
CC of increased stability, to be stable and to have low toxicity. The  
CC composition also comprises an agent which causes viral RNA to be  
CC inactive. The composition is used for preventing HCV replication in a  
CC system. The present sequence is encoded by a novel HCV cDNA sequence,  
CC which is used in the course of the invention

XX Sequence 2772 AA;

Query Match 99.5%; Score 3602; DB 3; Length 2772;  
Best Local Similarity 99.6%; Pred. No. 6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
DB 912 LAPITAYAOOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 971  
QY 61 GTRTIASPKGPVQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120  
DB 972 GTRTIASPKGPVQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 1031  
QY 121 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1032 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1091  
QY 181 RSPVFTDNSSPPVPOSFQVAHLHAPTGSKSTKVPAAQAQYKVLVLPNSVAATLGF 240  
DB 1092 RSPVFTDNSSPPVPOSFQVAHLHAPTGSKSTKVPAAQAQYKVLVLPNSVAATLGF 1151  
QY 241 AYMSKAHGIDPNIRTVGRTITTSPTITSTYTGKFLADGGCGGAYDIIICDECHSDTATS 300  
DB 1152 AYMSKAHGIDPNIRTVGRTITTSPTITSTYTGKFLADGGCGGAYDIIICDECHSDTATS 1211  
QY 301 ILGIGTVLDAQETAGARLVLTATPPGVSIVPHNIEEVALSTTGEIPFYGKAIPLEVI 360  
DB 1212 ILGIGTVLDAQETAGARLVLTATPPGVSIVPHNIEEVALSTTGEIPFYGKAIPLEVI 1271  
QY 361 KGRHLIFCHSKKCDLAAKLVALGINAVAYYRGLOVSVIPPIGDVVVATDALMTGYT 420  
DB 1272 KGRHLIFCHSKKCDLAAKLVALGINAVAYYRGLOVSVIPPIGDVVVATDALMTGYT 1331  
QY 421 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRTQRRORTGRGKPGIYRFVAPG 480  
DB 1332 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRTQRRORTGRGKPGIYRFVAPG 1391  
QY 481 ERPSGMFDSVLCYCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 540  
DB 1392 ERPSGMFDSVLCYCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGL 1451  
QY 541 THIDAHFLSOTKQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTPL 600  
DB 1452 THIDAHFLSOTKQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTPL 1511  
QY 601 YRLGAVQNEITLHPVTKYIMTCMSADLEVVSTWLVGGVLAALAYCLSTGCWVIIVGR 660  
DB 1512 YRLGAVQNEITLHPVTKYIMTCMSADLEVVSTWLVGGVLAALAYCLSTGCWVIIVGR 1571  
QY 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
DB 1572 VVLSGKPAIIPDREVLYREFDEMEEC 1597

RESULT 10  
ADN35976

ID ADN35976 standard; protein; 2772 AA.

XX AC ADN35976;

XX DT 17-JUN-2004 (first entry)

XX DE HCV cDNA clone #1 protein.

XX KW Antiviral; Vaccine; hepatitis C virus infection; HCV infection.

XX OS Hepatitis C virus.

XX PN EP1394255-A2.

XX PD 03-MAR-2004.

XX PF 16-MAR-1990; 2003EP-00016585.

XX PR 17-MAR-1989; 89US-00325338.

XX PR 20-APR-1989; 89US-00341334.

XX PR 18-MAY-1989; 89US-00355002.

XX PR 16-MAR-1990; 90EP-00302866.

XX PA (CHIR ) CHIRON CORP.

XX PI Houghton M, Choo Q, Kuo G;

XX DR WPI; 2004-193149/19.

XX DR N-PSDB; ADN35977.

XX PT Novel purified hepatitis C virus polypeptide comprising epitope encoded  
PT by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C  
PT virus.

XX PS Example 1; Fig 16; 79pp; English.

XX CC The present invention relates to hepatitis C virus (HCV) proteins and  
CC cDNA sequences. The sequences are useful in immunoassays for detecting  
CC antibodies directed against HCV antigen; preparing host cells transformed  
CC with a recombinant polynucleotide; screening antiviral agents and  
CC determining the effect of antiviral agent in inhibiting viral replication  
CC in cell culture system; and developing vaccine for treating HCV  
CC infection.

XX SQ Sequence 2772 AA;

Query Match 99.5%; Score 3602; DB 8; Length 2772;  
Best Local Similarity 99.6%; Pred. No. 6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
DB 912 LAPITAYAOOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 971  
QY 61 GTRTIASPKGPVQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120  
DB 972 GTRTIASPKGPVQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 1031  
QY 121 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1032 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1091  
QY 181 RSPVFTDNSSPPVPOSFQVAHLHAPTGSKSTKVPAAQAQYKVLVLPNSVAATLGF 240  
DB 1092 RSPVFTDNSSPPVPOSFQVAHLHAPTGSKSTKVPAAQAQYKVLVLPNSVAATLGF 1151  
QY 241 AYMSKAHGIDPNIRTVGRTITTSPTITSTYTGKFLADGGCGGAYDIIICDECHSDTATS 300  
DB 1152 AYMSKAHGIDPNIRTVGRTITTSPTITSTYTGKFLADGGCGGAYDIIICDECHSDTATS 1211

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 360  
Db 1212 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 1271  
Qy 361 KGGHLLIFCHSKKKDELAALVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYT 420  
Db 1272 KGGHLLIFCHSKKKDELAALVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYT 1331  
Qy 421 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 480  
Db 1332 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 1391  
Qy 481 ERPSGMFDDSVLCECYDAGCAWYELTTPAETTVRLRAYMNTFGLPVCODHLEFEGVFTGL 540  
Db 1392 ERPSGMFDDSVLCECYDAGCAWYELTTPAETTVRLRAYMNTFGLPVCODHLEFEGVFTGL 1451  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600  
Db 1452 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 1511  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWIVGR 660  
Db 1512 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWIVGR 1571  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 1572 VVLSGKPAIIPDREVLYREFDEMEEC 1597

## RESULT 11

AA14975  
ID AA14975 standard; protein; 2955 AA.

AC AA14975;

XX 20-MAR-2003 (revised)

DT 08-NOV-1999 (first entry)

XX Amino acid sequence of HCV-1 ORF.

XX Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;  
KW HCV infection; vaccine.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT Misc-difference 441

FT Misc-difference /note= "encoded by TT"

FT Misc-difference 461

FT Misc-difference /note= "encoded by CCCC"

XX EP939128-A2.

XX 01-SEP-1999.

XX 17-SEP-1990; 99EP-00101746.

XX 15-SEP-1989; 89US-00408045.

XX 21-DEC-1989; 89US-00456142.

XX 17-SEP-1990; 90EP-00310149.

XX (OYAA/) OYA A.

XX (CHIR ) CHIRON CORP.

XX Miyamura T, Saito I, Houghton M, Weiner AJ, Han J, Kolberg JA;

XX Cha T, Irvine BD;

XX WPI; 1999-480843/41.

XX N-PSDB; AA207656.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PS Disclosure; Fig 12; 132pp; English.

XX The invention provides two new isolates of hepatitis C virus (HCV), J1  
CC and J7. These two isolates comprise nucleotide and amino acid sequences  
CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences  
CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by  
CC hybridisation for diagnosis of NANBH infections. They may also be used to  
CC screen blood donors, donated blood and blood products for this infection.  
CC The isolates may also be used to isolate other naturally occurring  
CC variants of the virus. The polypeptides may be used as a vaccine for  
CC administration to patients to protect against infection with NANBH. The  
CC present sequence represents the amino acid sequence of HCV-1 ORF.  
CC (Updated on 20-MAR-2003 to correct PF field.) (Updated on 20-MAR-2003 to  
CC correct PR field.)  
XX

XX Sequence 2955 AA;

Query Match 99.5%; Score 3602; DB 2; Length 2955;

Best Local Similarity 99.6%; Pred. No. 66e-304;

Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQTRGLGCIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Db 1026 LAPITAYAAQTRGLGCIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy 61 GTRTIASPKGPVIOQMTNVDDLVGHPAPQGSRLTPCTCGSSDLVLTTHADVIPVRRR 120

Db 1086 GTRTIASPKGPVIOQMTNVDDLVGHPAPQGSRLTPCTCGSSDLVLTTHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRISYLLKSGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDIPVENLETTM 180

Db 1146 GDSRGSLLSPRISYLLKSGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVVPQSFQVAHLHAPGTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFG 240

Db 1206 RSPVFTDNSSPPVVPQSFQVAHLHAPGTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFG 1265

Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYGFADGGCGGAYDIIICDECHSTDATS 300

Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYGFADGGCGGAYDIIICDECHSTDATS 1325

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 360

Db 1326 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 1385

Qy 361 KGGHLLIFCHSKKKDELAALVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYT 420

Db 1386 KGGHLLIFCHSKKKDELAALVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYT 1445

Qy 421 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 480

Db 1446 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 1505

Qy 481 ERPSGMFDDSVLCECYDAGCAWYELTTPAETTVRLRAYMNTFGLPVCODHLEFEGVFTGL 540

Db 1506 ERPSGMFDDSVLCECYDAGCAWYELTTPAETTVRLRAYMNTFGLPVCODHLEFEGVFTGL 1565

Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 600

Db 1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPL 1625

Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWIVGR 660

Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWIVGR 1685

Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686

Db 1686 VVLSGKPAIIPDREVLYREFDEMEEC 1711

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

## RESULT 12

AA18541

ID AA18541 standard; protein; 2955 AA.

XX	AAB19541;
XX	AC
XX	DT
XX	15-JAN-2001 (first entry)
XX	Polyprotein encoded by sense strand of HCV.
DE	DE
XX	Hepatitis C virus; HCV; antisense polynucleotide; polyprotein; viral infectivity; viral replication.
KW	KW
OS	Hepatitis C virus.
XX	OS
PN	EP1034785-A2.
XX	XX
PD	13-SEP-2000.
XX	XX
PF	16-MAR-1990; 2000EP-00109602.
XX	XX
PR	17-MAR-1989; 89US-00325338.
PR	20-APR-1989; 89US-00341334.
PR	18-MAY-1989; 89US-00355002.
PR	16-MAR-1990; 90EP-00302866.
XX	XX
PA	(CHIR ) CHIRON CORP.
XX	XX
PI	Houghton M, Choo Q, Kuo G;
XX	XX
DR	WIPI; 2000-566891/53.
DR	N-PSDB; AAA75297.
XX	XX
PT	Novel composition comprising a hepatitis C virus antisense polynucleotide which is complementary to or corresponds to a sense strand of the virus genome, and selectively hybridizes to it.
PT	Example; Fig 17; 75pp; English.
XX	XX
CC	The specification describes a pharmaceutical composition which comprises a hepatitis C virus (HCV) antisense polynucleotide. The HCV is characterized by a positive stranded RNA genome which has 40% homology at the polypeptide level to a HCV polyprotein. The antisense polynucleotide binds to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability or chronicity. The antisense polynucleotides may also be designed to bind with high specificity, to be of increased stability, to be stable and to have low toxicity. The composition also comprises an agent which causes viral RNA to be inactive. The composition is used for preventing HCV replication in a system. The present sequence is encoded by a novel HCV cDNA sequence, which is used in the course of the invention
XX	XX
SQ	Sequence 2955 AA;
Query Match 99.5%; Score 3602; DB 3; Length 2955;	
Best Local Similarity 99.6%; Pred. No. 6.6e-304;	
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAAQTGRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 60 :
Dd	1026 LAPITAAQTGRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 1085
Qy	61 GTRTIASPKGVIQMYNTVDODLVGWPAPOGSRSLTPTCTCGSSDLYLVTTHADVIPVRRR 120
Dd	1086 GTRTIASPKGVIQMYNTVDODLVGWPAPOGSRSLTPTCTCGSSDLYLVTTHADVIPVRRR 1145
Qy	121 GDSRGSLSPRPISVLKSGSGGPLLCPAGHAAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180
Dd	1146 GDSRGSLSPRPISVLKSGSGGPLLCPAGHAAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
Qy	181 RSPVFTDNSSPPVPQSFQVAHLHAPTGSIGSKTKVPAAYAQQKYLVLPNSVAATLGFG 240
Dd	1206 RSPVFTDNSSPPVPQSFQVAHLHAPTGSIGSKTKVPAAYAQQKYLVLPNSVAATLGFG 1265
Qy	241 AYMSKAHGIDPNRTGVNFTITTGSPITVYSTVGKFLADGGCSGGAYDIIICDCHECHSTATS 300 :::

Db	1266	AYMSKAHGIDPNIRITGVRITTTGSPITYTYTGKFLADGSGGAYDIIILICDECHSIDATS	1326
Qy	301	ILGIGTVLQDAETAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLLEVI	360
Db	1326	ILGIGTVLQDAETAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLLEVI	1385
Qy	361	KGRHILIFCHSKKKKCDLAALKVALGINAVAYYRGDLVSVIPPIDGVVVVATDALMTGYT	420
Db	1386	KGRHILIFCHSKKKKCDLAALKVALGINAVAYYRGDLVSVIPSGDVVVVATDALMTGYT	1445
Qy	421	GDPSVIDNCNCTVQTVDPSLDPFTTETITLPODAVSRTOGRGTRGKPGIYRFVAPG	480
Db	1446	GDPSVIDNCNCTVQTVDPSLDPFTTETITLPODAVSRTOGRGTRGKPGIYRFVAPG	1505
Qy	481	ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFWEGVFTGL	540
Db	1506	ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFWEGVFTGL	1565
Qy	541	THIDAHFLSQTQSGENLPVLVAYQATVCARAQAPPPSDQMKCLIRLKP TLHGPTPL	600
Db	1566	THIDAHFLSQTQSGENLPVLVAYQATVCARAQAPPPSDQMKCLIRLKP TLHGPTPL	1625
Qy	601	YRLGAVQNEITLTHPVTKYIMTQMSADLEVVTTWLVGGVLAALAAAYCLSTGCWVIVGR	660
Db	1626	YRLGAVQNEITLTHPVTKYIMTQMSADLEVVTTWLVGGVLAALAAAYCLSTGCWVIVGR	1685
Qy	661	VWLSGRPAIIPDREVLVREFDEMEEC	686
Db	1686	VWLSGRPAIIPDREVLVREFDEMEEC	1711
RESULT 13			
ADN35978	ADN35978 standard; protein; 2955 AA.		
XX	AC	ADN35978;	
XX	DT	17-JUN-2004 (first entry)	
XX	DE	HCV cDNA clone #2 protein.	
XX	KW	Antiviral; Vaccine; hepatitis C virus infection; HCV infection.	
XX	OS	Hepatitis C virus.	
XX	PN	EP1394255-A2.	
XX	PD	03-MAR-2004.	
XX	PF	16-MAR-1990; 2003EP-00016585.	
XX	PR	17-MAR-1989; 89US-00325338.	
XX	PR	20-APR-1989; 89US-00341334.	
XX	PR	18-MAY-1989; 89US-00355002.	
XX	PR	16-MAR-1990; 90EP-00302866.	
XX	PA	(CHIR ) CHIRON CORP.	
XX	PI	Houghton M, Choo Q, Kuo G;	
XX	DR	WPI, 2004-193149/19.	
XX	DR	N-P5DB; ADN35979.	
XX	PT	Novel purified hepatitis C virus polypeptide comprising epitope encoded	
XX	PT	by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C	
XX	PT	virus.	
XX	PS	Example 1; Fig 17; 79pp; English.	
XX	CC	The present invention relates to hepatitis C virus (HCV) proteins and	
XX	CC	cDNA sequences. The sequences are useful in immunoassays for detecting	
XX	CC	antibodies directed against HCV antigen; preparing host cells transformed	
XX	CC	with a recombinant polynucleotide; screening antiviral agents and	

CC determining the effect of antiviral agent in inhibiting viral replication  
CC in cell culture system; and developing vaccine for treating HCV  
CC infection.

XX Sequence 2955 AA;

Query Match 99.5%; Score 3602; DB 8; Length 2955;  
Best Local Similarity 99.6%; Pred. No. 6.6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60  
Db :  
1026 LAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 1085  
Qy 61 GTRTIASPKGPVIOQMTNVDDLVGWPAPQGSRLTPTCTCGSSDLXLVTRHADVIPVRRR 120  
Db :  
1086 GTRTIASPKGPVIOQMTNVDDLVGWPAPQGSRLTPTCTCGSSDLXLVTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db :  
1146 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 240  
Db :  
1206 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300  
Db :  
1266 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDDAETAGARLVVLATAPPGSVTVPHNTEEEVALSTTGSIPIYKGAIPLEVI 360  
Db :  
1326 ILGIGTVLDDAETAGARLVVLATAPPGSVTVPHNTEEEVALSTTGSIPIYKGAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKCDLAAKLVAGINAVAYRGLDVSIVPTPGDVAATLMTGYT 420  
Db :  
1386 KGRHLIFCHSKKCDLAAKLVAGINAVAYRGLDVSIVPTPGDVAATLMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRRGTRGKPGIYRFVAPG 480  
Db :  
1446 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRRGTRGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCBCEYDAGCAWYELTPAETTVRLRAYMNTGVLPCVQDHLFEFEGVPTGL 540  
Db :  
1506 ERPSGMFDSVLCBCEYDAGCAWYELTPAETTVRLRAYMNTGVLPCVQDHLFEFEGVPTGL 1565  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMWKCLIRLXPTLHGPTPL 600  
Db :  
1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMWKCLIRLXPTLHGPTPL 1625  
Qy 601 YRLGAVONEITLTHPVTKYIMTMSADLEVVTSTWLVGVLAAALAYCLSTGCWVIAGR 660  
Db :  
1626 YRLGAVONEITLTHPVTKYIMTMSADLEVVTSTWLVGVLAAALAYCLSTGCWVIAGR 1685  
Qy 661 VVLGKPAIIPDREVLYREFDEMEEC 686  
Db :  
1686 VVLGKPAIIPDREVLYREFDEMEEC 1711

RESULT 14

AAR90931

ID AAR90931 standard; protein; 3011 AA.

XX AAR90931;

AC AAR90931;

DT 25-MAR-2003 (revised)

DT 15-MAY-1996 (first entry)

XX Hepatitis C virus polyprotein.

XX Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection; diagnosis;

XX antibodies.

XX

OS Hepatitis C virus.

XX Key Location/Qualifiers

FT Misc-difference 1..122

FT /label= antigen

FT /note= "C22; AAR90936"

FT Misc-difference 199..328

FT /label= antigen

FT /note= "S2; AAR90935"

FT Misc-difference 1192..1457

FT /label= antigen

FT /note= "C33C; AAR90932"

FT Misc-difference 1569..1931

FT /label= antigen

FT /note= "C100; AAR90933"

FT Misc-difference 2054..2464

FT /label= antigen

FT /note= "NS5; AAR90934"

FT

XX EP693687-A1.

PN

XX 24-JAN-1996.

PD

XX 03-APR-1991; 95EP-00114016.

PF

XX 04-APR-1990; 90US-00504352.

PR

XX (CHIR ) CHIRON CORP.

XX

XX Houghton M, Choo Q, Kuo G;

PI

XX WPI; 1996-117956/13.

DR

XX N-PSDB; AAT12710.

XX

XX Combinations of synthetic Hepatitis C Virus antigens - provide more

PT effective diagnosis of Non-A, Non-B Hepatitis.

XX

XX Disclosure; Fig 1(A-Y); 53pp; English.

PS

XX The combination comprises an HCV antigen from the C domain (pref. C22 -

CC AAR90936) and at least one HCV antigen from the NS3 (pref. C33C -

CC AAR90932), NS4 (pref. C100 - AAR90933), S (pref. S2 - AAR90935) or NS5

CC (AAR90934) domain. The antigens may in the form of a fusion protein, a

CC simple physical mixture, or the individual antigens commonly bound to a

CC solid matrix. They are pref. prepd. by recombinant DNA techniques

CC (primers are given in AAT12711-T12716), but can be synthesised or

CC isolated from HCV using affinity chromatography. (Updated on 25-MAR-2003

CC to correct PF field.)

XX

XX Sequence 3011 AA;

Query Match 99.5%; Score 3602; DB 2; Length 3011;

Best Local Similarity 99.6%; Pred. No. 6.6e-304;

Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60

Db :  
1026 LAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 1085

Qy 61 GTRTIASPKGPVIOQMTNVDDLVGWPAPQGSRLTPTCTCGSSDLXLVTRHADVIPVRRR 120

Db :  
1086 GTRTIASPKGPVIOQMTNVDDLVGWPAPQGSRLTPTCTCGSSDLXLVTRHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180

Db :  
1146 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 240

Db :  
1206 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 1265

Qy 241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300

XX



Db	1266	AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKGLADGCGGAVDIIICDECHSTDATS	1325
Qy	301	ILIGITVLDQAEATAGARLVVLATATPPGVTWPHNIEEVALSTTGEIIFYGKAIPLEVI	360
Db	1326	ILIGITVLDQAEATAGARLVVLATATPPGVTWPHNIEEVALSTTGEIIFYGKAIPLEVI	1385
Qy	361	KGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSIPIIGDVVVVATDALTMTGYT	420
Db	1386	KGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSIPIIGDVVVVATDALTMTGYT	1445
Qy	421	GFDSVIDCNTCTVTQVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPG	480
Db	1446	GFDSVIDCNTCTVTQVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPG	1505
Qy	481	ERPSGMFSDSVLCEYDAGCAYVELTPAETTVRLRAYMNTPLPVQCDHLEFWEGVFTGL	540
Db	1506	ERPSGMFSDSVLCEYDAGCAYVELTPAETTVRLRAYMNTPLPVQCDHLEFWEGVFTGL	1565
Qy	541	THIDAHFLSQTQSGENLPVLVAYQATVCARAQPPSDQMWKCLIRLKLHGTPTLL	600
Db	1566	THIDAHFLSQTQSGENLPVLVAYQATVCARAQPPSDQMWKCLIRLKLHGTPTLL	1625
Qy	601	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGR	660
Db	1626	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGR	1685
Qy	661	VVLGSKPAIIPDREVLRYRFEDEMEEC	686
Db	1686	VVLGSKPAIIPDREVLRYRFEDEMEEC	1711
RESULT 15			
AAW34480	ID	AAW34480 standard; protein; 3011 AA.	
XX	AC	AAW34480;	
XX	AC	AAW34480;	
XX	AC	AAW34480;	
DT	25-MAR-2003	(revised)	
DT	16-MAR-1998	(first entry)	
DE	XX	HCV polyprotein.	
XX	XX	PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;	
KW	KW	C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;	
KW	KW	NS4.	
XX	XX	Hepatitis C virus.	
OS	XX		
FH	Key	Location/Qualifiers	
FT	Misc-difference	366 /note= "can optionally be Arg"	
FT	Misc-difference	372 /note= "can optionally be Thr"	
FT	Misc-difference	867 /note= "can optionally be Thr"	
FT	Misc-difference	1341 /note= "can optionally be Val"	
FT	Misc-difference	2148 /note= "can optionally be Ile"	
FT	Misc-difference	2883 /note= "can optionally be Asn"	
FT	Misc-difference	3681 /note= "can optionally be Ser"	
FT	Misc-difference	3690 /note= "can optionally be Thr"	
FT	Misc-difference	4167 /note= "can optionally be Leu"	
FT	Misc-difference	4323 /note= "can optionally be Val"	
FT	Misc-difference	4701 /note= "can optionally be Tyr"	
FT	Misc-difference	4752 /note= "can optionally be Ser"	

FT	Misc-difference	5970 /note= "can optionally be Gly"	
FT	Misc-difference	6183 /note= "can optionally be His"	
FT	Misc-difference	6186 /note= "can optionally be Cys"	
FT	Misc-difference	6402 /note= "can optionally be Val"	
FT	Misc-difference	7386 /note= "can optionally be Ser"	
FT	Misc-difference	7494 /note= "can optionally be Phe"	
FT	Misc-difference	7497 /note= "can optionally be Ala"	
FT	Misc-difference	7845 /note= "can optionally be Phe"	
FT	Misc-difference	8409 /note= "can optionally be Gly"	
FT	Misc-difference	9102 /note= "can optionally be Gly"	
FT	Misc-difference	9327 /note= "can optionally be Pro"	
XX	US5683864-A.		
PN	XX	04-NOV-1997.	
XX	XX	07-JUL-1992; 92US-00910760.	
XX	XX	18-NOV-1987; 87US-00122714.	
PR	30-DEC-1987; 87US-00139886.		
PR	26-FEB-1988; 88US-00161072.		
PR	06-MAY-1988; 88US-00191263.		
PR	26-OCT-1988; 88US-00263584.		
PR	14-NOV-1988; 88US-00271450.		
PR	17-MAR-1989; 89US-00325338.		
PR	20-APR-1989; 89US-00341334.		
PR	21-APR-1989; 89US-00353896.		
PR	18-MAY-1989; 89US-00355002.		
PR	04-APR-1990; 90US-00504352.		
XX	(CHIR ) CHIRON CORP.		
XX	Kuo G, Houghton M, Choo Q;		
XX	WPI; 1997-548976/50.		
DR	N-PSDB; AAT99981.		
XX	Combination of three hepatitis C virus antigens - used for detection of specific antibodies to diagnose infection.		
PT	Disclosure; Col 25-46; 57pp; English.		
XX	This sequence represents the Hepatitis C virus polyprotein. Fragments of the DNA encoding this sequence can be amplified and used in the combination of HCV antigens of the invention. The HCV antigen combination comprises an antigen (Ag1) comprising the C domain (i.e. amino acids (aa) 1-120 of the HCV polyprotein), or its immunologically reactive fragment containing at least 8 aa. It also comprises two additional antigens from two different polyprotein domains, including at least 8 aa from the NS3, NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein. Alternatively, Ag1 contains at least 8 aa from the 1-122 or 9-177 aa regions of the HCV polyprotein. These antigen combinations are used diagnostically to detect anti-HCV antibodies, using any standard immunoassay format. These antigen combinations have a broader range of reactivity with antibodies than any antigen individually. (Updated on 25-MAR-2003 to correct PR field.)		
XX	Sequence 3011 AA;		
SQ	Query Match	99.5%; Score 3602; DB 2; Length 3011;	
	Best Local Similarity	99.6%; Pred. No. 6.8e-304;	

Matches 683;		Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	1	MAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60							
Db	1026	LAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085							
Qy	61	GTRTIASPKGPVIOYMYTNVDODLVGWPAPOGSRSLTCTCGSSDLYLVTTRHADVIPVRRR 120							
Db	1086	GTRTIASPKGPVIOYMYTNVDODLVGWPAPOGSRSLTCTCGSSDLYLVTTRHADVIPVRRR 1145							
Qy	121	GDSRGSLLSPRPISYLAGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180							
Db	1146	GDSRGSLLSPRPISYLAGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205							
Qy	181	RSPVFTDNSPPVPVQSFQVAHLHAPTGSKGSTKVPAAAYAAQGYKVLVLPNPSVAATLGFG 240							
Db	1206	RSPVFTDNSPPVPVQSFQVAHLHAPTGSKGSTKVPAAAYAAQGYKVLVLPNPSVAATLGFG 1265							
Qy	241	AYMSKAHGIDPNIRHTGVRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSTDATS 300							
Db	1266	AYMSKAHGIDPNIRHTGVRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSTDATS 1325							
Qy	301	ILGIGTVLDQAEATAGARLVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAI PLEVI 360							
Db	1326	ILGIGTVLDQAEATAGARLVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAI PLEVI 1385							
Qy	361	KGRHLIFCHSKKCCDELAALKVALGINAVAYRGLDVSVIPPIGDVVVVATDALMTGYT 420							
Db	1386	KGRHLIFCHSKKCCDELAALKVALGINAVAYRGLDVSVIPPTSGDVVVVATDALMTGYT 1445							
Qy	421	GDFDSVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRRTORRGRTGRKPGIYRFVAPG 480							
Db	1446	GDFDSVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRRTORRGRTGRKPGIYRFVAPG 1505							
Qy	481	ERPSGMFDSVLCFCYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 540							
Db	1506	ERPSGMFDSVLCFCYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 1565							
Qy	541	THIDAHFLSQTQSGENLPYLVAQATVCARAAQAPPPSWDQMKCLIRLKPTRLHGPTPLL 600							
Db	1566	THIDAHFLSQTQSGENLPYLVAQATVCARAAQAPPPSWDQMKCLIRLKPTRLHGPTPLL 1625							
Qy	601	YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVI VGR 660							
Db	1626	YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVI VGR 1685							
Qy	661	VVLSGKPAIIPDREVLYREDEMEEC 686							
Db	1686	VVLSGKPAIIPDREVLYREDEMEEC 1711							

Search completed: November 7, 2005, 20:09:55  
Job time : 84.2548 secs

1960

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:00:21 ; Search time 14.1714 Seconds  
(without alignments)  
4657.604 Million cell updates/sec

Title: US-10-658-782-2  
Perfect score: 3619  
Sequence: 1 MAPITAYAAQOTRGLGCIIT.....PAIIPREVLYREFDEMEEC 686

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3602	99.5	3011	1 GNVWC3	genome polyprotein
2	3542	97.9	3011	1 S40770	genome polyprotein
3	3513	97.1	3011	1 GNVVTC	genome polyprotein
4	3426	94.7	3010	1 GNVVTC	genome polyprotein
5	3417	94.4	3010	1 A45573	genome polyprotein
6	3408	94.2	3010	1 GNVWCJ	genome polyprotein
7	3402	94.0	3010	1 GNVVTM	genome polyprotein
8	3375	93.3	3010	1 S18030	genome polyprotein
9	3082	85.2	3014	1 JC5620	genome polyprotein
10	2989	82.6	3033	1 JQ1303	genome polyprotein
11	2979	82.3	3033	1 GNVVJ8	genome polyprotein
12	1993	55.1	876	2 PC2219	polyprotein - hepa
13	1972.5	54.5	286	2 S68016	ATPase/RNA helicase
14	1388	38.4	492	2 PS0326	polyprotein - hepa
15	1383	38.2	716	2 JQ1366	polyprotein - hepa
16	1101	30.4	3005	2 T08841	polyprotein - dour
17	1075	29.7	2970	2 T08839	polyprotein - marm
18	1055	29.2	216	2 S21337	genome polyprotein
19	1001	27.7	194	2 S06067	nonstructural prot
20	998	27.6	194	2 A54317	probable nonstruct
21	980	27.1	182	2 S32748	genome polyprotein
22	947	26.2	184	2 A61196	genome polyprotein
23	728	20.1	209	2 PC1306	genome polyprotein
24	720	19.9	135	2 PS0327	polyprotein - hepa
25	719	19.9	135	2 PS0328	polyprotein - hepa
26	693	19.1	135	2 PC0329	polyprotein - hepa
27	397	11.0	142	2 PC1307	genome polyprotein
28	310	8.6	102	2 PC6028	genome polyprotein
29	280	7.7	132	2 PQ0394	genome polyprotein

30	280	7.7	132	2	PQ0396	genome polyprotein
31	277	7.7	157	2	PQ0401	genome polyprotein
32	276	7.6	125	2	S35629	hypothetical prote
33	273	7.5	3898	1	GNWVHC	genome polyprotein
34	273	7.5	3898	1	GNWVHB	genome polyprotein
35	271	7.5	3898	2	S57437	genome polyprotein
36	268.5	7.4	3968	1	GNWVEV	genome polyprotein
37	267.5	7.4	3898	1	A44217	genome polyprotein
38	265.5	7.3	266	2	PQ0393	polyprotein - hog
39	264	7.3	3898	2	S58295	genome polyprotein
40	227.5	6.3	3341	1	A42996	genome polyprotein
41	223	6.2	3434	1	GNWVWV	genome polyprotein
42	202	5.6	3125	1	GNVSPB	genome polyprotein
43	202	5.6	3140	2	S47508	genome polyprotein
44	201	5.6	3140	1	GNVSRA	genome polyprotein
45	201	5.6	3141	1	GNVSPD	genome polyprotein

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)  
N:contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A39166; PQ0403; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; MUID:91172826; PMID:1848704  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:Cross-references: UNIPROT:P26664; GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0403  
A:Molecule type: Genomic RNA  
A:Residues: 1577-1633 <CHA>  
A:Cross-references: DDBJ:D10128  
A:Experimental source: isolates E-b16  
A:Accession: PQ0404  
A>Status: Preliminary  
A:Molecule type: Genomic RNA  
A:Residues: 1577-1633 <CH2>  
A:Experimental source: isolates E-b17  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F1-115/Product: capsid protein C #status predicted <CPC>  
F116-191/Product: envelope protein M #status predicted <EPM>  
F192-389/Product: major envelope protein E #status predicted <MEB>  
F390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F1007-1615/Product: hepatitisvirin #status predicted <NS3>  
F1230-1237/Region: nucleotide-binding motif A (P-loop)  
F1312-1317/Region: nucleotide-binding motif B  
F1316-1319/Region: DEXH motif  
F1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 99.5%; Score 3602; DB 1; Length 3011;  
Best Local Similarity 99.6%; Pred. No. 7.4e-240;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAAQOTRGLGCIITSTGRDKNOVEGVQIVSTAQTFLATCINGVCTVYHCA 60  
:|||||

Db 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSSGGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSSGGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDSSPPVQSFQVAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDSSPPVQSFQVAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDQAETAGARLVWLATATPPGSVTVPHNIEEVALSTTGIBPFYKGAIPLEVI 360  
Db 1326 ILGIGTVLDQAETAGARLVWLATATPPGSVTVPHNIEEVALSTTGIBPFYKGAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDPSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDPSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 600  
Db 1566 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAYCLSTGCWVI 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAYCLSTGCWVI 1685  
Qy 661 VLSGKPAIIPDREVLRYREFDEMEEC 686  
Db 1686 VLSGKPAIIPDREVLRYREFDEMEEC 1711  
RESULT 2  
S40770  
genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: S40770; PC1285  
R;Okamoto, H.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S40770  
A;Accession: S40770  
A;Molecule type: genomic RNA  
A;Residues: 1-3011 <OKA>  
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A;Reference number: PC1284; MUID:91013116; PMID:2170712  
A;Accession: PC1285  
A;Molecule type: genomic RNA  
A;Residues: 1-513 <OK2>  
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512  
A;Experimental source: isolate HC-J1  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <BPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1065/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
Query Match 97.9%; Score 3542; DB 1; Length 3011;  
Best Local Similarity 96.9%; Pred. No. 1e-235;  
Matches 665; Conservative 10; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 MAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSSGGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSSGGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDSSPPVQSFQVAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDSSPPVQSFQVAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDQAETAGARLVWLATATPPGSVTVPHNIEEVALSTTGIBPFYKGAIPLEVI 360  
Db 1326 ILGIGTVLDQAETAGARLVWLATATPPGSVTVPHNIEEVALSTTGIBPFYKGAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDPSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDPSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 600  
Db 1566 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAYCLSTGCWVI 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAYCLSTGCWVI 1685  
Qy 661 VLSGKPAIIPDREVLRYREFDEMEEC 686  
Db 1686 VLSGKPAIIPDREVLRYREFDEMEEC 1711  
RESULT 3  
GNVWCH  
genome polyprotein - hepatitis C virus (strain H)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C;Accession: A36814; A41546  
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
submitted to GenBank, July 1992  
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992  
A;Reference number: A36814  
A;Accession: A36814  
A;Molecule type: Genomic RNA  
A;Residues: 1-3011 <INC>  
A;Cross-references: UNIPROT:P27958; GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992  
A;Reference number: A41546; MUID:92052256; PMID:1658800  
A;Contents: annotation  
A;Note: neither amino acid nor nucleotide sequence is given  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
F;1-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein  
F;1230-1237/Product: hepatitis C virus genome polyprotein  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 97.1%; Score 3513; DB 1; Length 3011;  
Best Local Similarity 96.6%; Pred. No. 1e-233;  
Matches 663; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 60  
Db 1026 LAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 1085

Qy 61 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 180  
Db 1146 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 1205

Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 1265

Qy 241 AYMSKAHGIDPNIRGVRTITTSPTTYSTYKFLADGCGSGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRGVRTITTSPTTYSTYKFLADGCGSGAYDIIICDECHSTDATS 1325

Qy 301 ILIGTGLVLDQAETAGARLVWLATATPPGVSVPVPHNIEEVALSTTGEIPFYKAIPL 360  
Db 1326 ILIGTGLVLDQAETAGARLVWLATATPPGVSVPVPHNIEEVALSTTGEIPFYKAIPL 1385

Qy 361 KGRHLIFCHSKKKCKDELAALVALGINAVAYRGLDVSVPPIGDVVVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKCKDELAALVALGINAVAYRGLDVSVPPIGDVVVVVATDALTMTGYT 1445

Qy 421 GFDSVIDQNTCTVQTFDPSLDPTTETITILPQAVSRTQGRGKPGIYRFVAPG 480  
Db 1446 GFDSVIDQNTCTVQTFDPSLDPTTETITILPQAVSRTQGRGKPGIYRFVAPG 1505

Qy 481 BRPSGMFDSVLCCEYDAGCAWELTPAETTVRLRAYMNTPLVCQDHLFEWGVFTGL 540  
Db 1506 BRPSGMFDSVLCCEYDAGCAWELTPAETTVRLRAYMNTPLVCQDHLFEWGVFTGL 1565

Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKCLIRLKTPLHGPTPL 600

Db 1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKCLIRLKTPLHGPTPL 1625

Qy 601 YRLGAVQNEITLHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVIYGR 660

Db 1626 YRLGAVQNEITLHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVIYGR 1685

Qy 661 VVLSGKPAIIPDRVLYREFDEMEEC 686

Db 1686 IVLSGKPAIIPDRVLYREFDEMEEC 1711

RESULT 4  
GNMVT  
genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: A38465  
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
A;Title: Structure and organization of the hepatitis C virus genome isolated from human  
A;Reference number: A38465; MUID:91140698; PMID:1847440  
A;Accession: A38465  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <TAK>  
A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G32977  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
F;1-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 94.7%; Score 3426; DB 1; Length 3010;  
Best Local Similarity 92.0%; Pred. No. 1e-227;  
Matches 631; Conservative 37; Mismatches 18; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 60  
Db 1026 LAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 1085

Qy 61 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GSKTLAAPKGPITQYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 180  
Db 1146 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETMM 1205

Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 1265

Qy 241 AYMSKAHGIDPNIRGVRTITTSPTTYSTYKFLADGCGSGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRGVRTITTSPTTYSTYKFLADGCGSGAYDIIICDECHSTDATS 1325

Qy 301 ILIGTGLVLDQAETAGARLVWLATATPPGVSVPVPHNIEEVALSTTGEIPFYKAIPL 360  
Db 1326 ILIGTGLVLDQAETAGARLVWLATATPPGVSVPVPHNIEEVALSTTGEIPFYKAIPIEAI 1385

Qy 361 KGRHLIFCHSKKKCKDELAALVALGINAVAYRGLDVSVPPIGDVVVVVATDALTMTGYT 420

Db 1386 RGRHLLFCHSKKCDLAALKSLGLGINAVAYRGLDVSVPITIGDVVVATDALMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVQTVDVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVAPG 480  
Db 1446 GDFDSVIDCNTCTVQTVDVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVTPG 1505  
Qy 481 ERPSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPTVKYIMTCHMSADLEVVTSTWLVGGVLAALAYCYLSTGCWVIWGR 660  
Db 1626 YRLGAVQNEITLTHPTVKYIMTCHMSADLEVVTSTWLVGGVLAALAYCYLSTGWSWIVGR 1685  
Qy 661 VVLGSKPAIIPDREVLRYREDEMEEC 686  
Db 1686 IILSGRPAIVPDRELLYQSFDEMEEC 1711

RESULT 5  
A45573  
genome polyprotein - hepatitis C virus (strain JT)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome from a single Japanese carrier: s  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: A45573  
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
Virus Res. 23, 39-53, 1992  
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s  
A;Reference number: A45573; MUID:92295714; PMID:1318627  
A;Accession: A45573  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-3010 <TAN>  
A;Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1;  
A;Experimental source: HCV-JT  
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome from a single Japanese patient's  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 94.4%; Score 3417; DB 1; Length 3010;  
Best Local Similarity 92.0%; Pred. No. 4.3e-217;  
Matches 631; Conservative 34; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MAPTAAQOTRGLGCIITSLGRDNQVEGEVQIVSTAQTPLATCINGCVCHTVVHGA 60  
Db 1026 LAPITAAQOTRGLGCIITSLGRDNQVEGEVQIVSTAQTPLATCINGCVCHTVVHGA 1085  
Qy 61 GTRTIASPKGPIQMTYNTVDQDLGVWPAQSGRSRLTPTCTGSSDLVLTVRHADVIPVRRR 120  
Db 1086 GSXTLAGPKGPITQMTYNTVDQDLGVWPAQSGRSRLTPTCTGSSDLVLTVRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGSGGGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDGRGSLLSPRPISYLGSGGGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVPOSFQVAHLHAPGTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPOSFQVAHLHAPGTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGIDPNIRTTGSPITYSTYKFLADGGCGSGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTTGSPITYSTYKFLADGGCGSGGAYDIIICDECHSTDTST 1325  
Qy 301 ILGITVLDQAEATAGARLVVLTATPPGSVTVPHNIEEVALSTTGTPIPYGKAIPLEVI 360  
Db 1326 ILGITVLDQAEATAGARLVVLTATPPGSVTVPHNIEEVALSTTGTPIPYGKAIPLEAI 1385  
Qy 361 KGRHLLFCHSKKCDLAALKSLGLGINAVAYRGLDVSVPITIGDVVVATDALMTGYT 420  
Db 1386 KGRHLLFCHSKKCDLAALKSLGLGINAVAYRGLDVSVPITIGDVVVATDALMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVQTVDVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVAPG 480  
Db 1446 GDFDSVIDCNTCTVQTVDVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVTPG 1505  
Qy 481 ERPSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPTVKYIMTCHMSADLEVVTSTWLVGGVLAALAYCYLSTGCWVIWGR 660  
Db 1626 YRLGAVQNEITLTHPTVKYIMTCHMSADLEVVTSTWLVGGVLAALAYCYLSTGWSWIVGR 1685  
Qy 661 VVLGSKPAIIPDREVLRYREDEMEEC 686  
Db 1686 IILSGRPAIVPDRELLYQSFDEMEEC 1711

RESULT 6  
GNWVCJ  
genome polyprotein - hepatitis C virus (strain J)  
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: A39253; PS0086  
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A;Reference number: A39253; MUID:91088550; PMID:2175903  
A;Accession: A39253  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <KAT>  
A;Cross-references: UNIPROT:P26662; GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611  
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence varia  
A;Reference number: PS0086  
A;Accession: PS0086  
A;Molecule type: genomic RNA  
A;Residues: 2650-2707 <KAT>  
A;Experimental source: Japanese isolate  
C;Comment: The cleavage sites of this polyprotein have not been determined.  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome from Japanese patients  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B



F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2  
Query Match 94.2%; Score 3408; DB 1; Length 3010;  
Best Local Similarity 91.4%; Pred. No. 1.8e-226;  
Matches 627; Conservative 38; Mismatches 21; Indels 0; Gaps 0;  
Qy 1 MAPITAAQOTRGLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Db 1026 LAPITAAQOTRGLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 1085  
Qy 61 GTRTIASPKGPVIQMTYNVDQDLVGPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPVRR 120  
Db 1086 GSKTLAGPKGPIQMTYNVDQDLVGPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPVRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVPOQFOVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPOQFOVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDOAETAGARLVLATATPPGTVTPHPNIEEVALSTGEIPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDOAETAGARLVLATATPPGTVTPHPNIEEVALSTGEIPFYGKAIPLEVI 1385  
Qy 361 KGBRHLLFCHSKKKDELAALKVALGINAVAYRGLDVSVIPIGDDVVVATDALTGT 420  
Db 1386 KGBRHLLFCHSKKKDELAALKVALGINAVAYRGLDVSVIPIGDDVVVATDALTGT 1445  
Qy 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITTLPODAVSRTQRRGRTGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITTLPODAVSRTQRRGRTGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQTQKAGNDLPYLVAQVTCARAQAPPSPSDQMWKCLIRLKLPTLHGPTLL 600  
Db 1566 THIDAHFLSQTQKAGNDLPYLVAQVTCARAQAPPSPSDQMWKCLIRLKLPTLHGPTLL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 1685  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 1686 IILSGKPAVVDREVLYQEFDEMEEC 1711

RESULT 7  
GNVVTW  
Genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A40244  
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992  
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A:Reference number: A40244; MUID:92230206; PMID:1314449  
A:Accession: A40244

A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: UNIPROT:P29846; GB:M84754  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F;1-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 207  
Query Match 94.0%; Score 3402; DB 1; Length 3010;  
Best Local Similarity 91.4%; Pred. No. 4.7e-226;  
Matches 627; Conservative 38; Mismatches 21; Indels 0; Gaps 0;  
Qy 1 MAPITAAQOTRGLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Db 1026 LAPITAAQOTRGLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 1085  
Qy 61 GTRTIASPKGPVIQMTYNVDQDLVGPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPVRR 120  
Db 1086 GSKTLAGPKGPIQMTYNVDQDLVGPAPQSGRSLSLTCTCGSSDLYLVTRHADVIPVRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVPOQFOVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPOQFOVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDOAETAGARLVLATATPPGTVTPHPNIEEVALSTGEIPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDOAETAGARLVLATATPPGTVTPHPNIEEVALSTGEIPFYGKAIPLEVI 1385  
Qy 361 KGBRHLLFCHSKKKDELAALKVALGINAVAYRGLDVSVIPIGDDVVVATDALTGT 420  
Db 1386 KGBRHLLFCHSKKKDELAALKVALGINAVAYRGLDVSVIPIGDDVVVATDALTGT 1445  
Qy 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITTLPODAVSRTQRRGRTGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITTLPODAVSRTQRRGRTGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFSSVLCBCYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQTQKAGNDLPYLVAQVTCARAQAPPSPSDQMWKCLIRLKLPTLHGPTLL 600  
Db 1566 THIDAHFLSQTQKAGNDLPYLVAQVTCARAQAPPSPSDQMWKCLIRLKLPTLHGPTLL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 1685  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 1686 IILSGKPAVVDREVLYQEFDEMEEC 1711

S18030  
Genome polyprotein - hepatitis C virus (isolate JX1)  
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Variety: isolate JX1  
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: S18030; S33570; A48332; S18029  
R:Honda, M.; Kaneko, S.; Maehashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
A:Reference number: S18028  
A:Accession: S18030  
A:Molecule type: Genomic RNA  
A:Residues: 1-3010 <H>  
A:Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59478  
A:Experimental source: Isolate JX1 from an individual  
R:Honda, M.; Kaneko, S.; Uncura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A>Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A:Reference number: A48332; MUID:93119270; PMID:8380322  
A:Accession: S33570  
A:Molecule type: Genomic RNA  
A:Residues: 1-547,'T','549-621','V','623-624','S','626-652','DL','655-761','T','763-782 <H>  
A:Cross-references: EMBL:X61591  
A>Note: This sequence is inconsistent with the nucleotide translation  
A>Note: The authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
A>Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIPI:121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F:2-115/Product: capsid protein C #status Predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepacivirin #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

	Query Match	93.3%	Score 3375;	DB 1;	Length 3010;
	Best local Similarity	91.3%	Pred. No. 3.4e-24;		
	Matches 626;	Conservative 33;	Mismatches 27;	Indels 0;	Gaps 0;
Qy	1	MAPTAYAAQTGRGLGCIITSLTGRDNKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGA	60	:	:
Dd	1026	LAPITAVSQTRGLFGCIIVTSLTGRDNKQVEGEAOVVSTATQSFATCVGVCTVYHGA	1085	:	:
Qy	61	GTRIASPQGPVIOMYNVDODLVGWPAQGSRSLSLTCTCGSSDLYLVTHRADVIPVRR	120	:	:
Dd	1086	GSKTLAGEPGINQMNYNDQDLVGWMQAPSGAASLTPTCYGSSDLYLVTHRADVIPVRR	1145	:	:
Qy	121	GDSRGSLSPRPISYLKGSSGGPLLCPACHAVGIFRAAVCTRGVAKAVDFIPVENLETTM	180	:	:
Dd	1146	GDSRGSLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM	1205	:	:
Qy	181	RSPVFTDNSSPPVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLPNSVAATLGFG	240	:	:
Dd	1206	RSPVFTDNSSPPVAPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLPNSVAATLGFG	1265	:	:
Qy	241	AYMSKAHGIDNIRGVRTIITGSPITYTYTKGLADGCCGGAYDIILICDECHSTDATS	300	:	:
Dd	1266	AYMSKAHGVNDNISIGVRTIITGAPITYTYTKGLADGCCGGAYDIILICDECHSTDSTS	1325	:	:
Qy	301	ILIGTGLVLDQAETAGARLWLATTPGGSVTVPHPNIIEVALSTTGEPFYGKAIPLEVI	360	:	:
Dd	1326	ILIGTGLVLDQAETAGARLWLAAATPPGGSVTVPHPNIEVALPNTGEIPFYGKAIPLETI	1385	:	:

```
Db 1147 GDTASLLSPRIISYLGSSGGPIMCPSPGHVGVFRAAVCTRGVAKALEFVFPVENLETTM 1206
Qy 181 RSPVFTNSPPVPOQSFQVLAHLPATGSGKSTKVPAAQYKVLVLPNSVAATLGF 240
Db 1207 RSPVFTNSPPVPOQSFQVLAHLPATGSGKSTKVPAAQYKVLVLPNSVAATLGF 1266
Qy 241 AYMSKAHGIDPNIRTVGRTITGSPITYSTYKFLADGGCGGGAYDIIICDECHSDATS 300
Db 1267 AYMSRAYGVDPNIRTVGRTITGAGITYSTYKGFADGGCGGGAYDIIICDECHSDATS 1326
Qy 301 ILGIGTVDQAEAGARLVLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db 1327 ILGIGTVDQAEAGARLVLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 1386
Qy 361 KGGRHLLFCHSKKCKDELAALVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYT 420
Db 1387 KGGRHLLFANOKKKAETAKONKPGKAVAYRGLDVAVIPATGDVVVSTDALMTGFT 1446
Qy 421 GDFSDVDCNTCVTQTVDVDFSLDPTFTTITLPODASVRSRRGRTGRGKPGIYRFVAPG 480
Db 1447 GDFSDVDCNSAVTQTVDFSLDPTFTTITLPODASVRSRRGRTGRGKPGIYRYVSSG 1506
Qy 481 ERPSGMFSDSVLCYDAGCAWYELTAEITTVRLRAYMNTPGLPVCDHLEFWEVGTGL 540
Db 1507 ERPSGIFSDSVVLCYDAGCAWYDLTAEITTVRLRAYLNTPLGLPVCDHLEFWEVGTGL 1566
Qy 541 THIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDQMWKCLIRLKLPTLHGPTPL 600
Db 1567 TNIDAHMLSQKQGENFPYLVAQVATVCARAKAPPPSDQMWKCLIRLKLPTLHGPTPL 1626
Qy 601 YRLGAVQNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
Db 1627 YRLGAVQNEITLTHPTVKYIMACNSADLEVTSTWLVGGVLAALAAAYCLTVGSVAIVGR 1686
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686
Db 1687 IILSGREPAITPDREVLYQDFDEMEEC 1712

RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: JQ1303
R:Okamoto, H.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y. J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A:Reference number: JQ1303; MUID:92044440; PMID:1658196
A:Accession: JQ1303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: UNIPROT:P26660; GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A:Experimental source: isolate HC-J6 from a Japanese individual
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1867-2017/Product: nonstructural protein NS4a #status predicted <N4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <N05>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 82.6%; Score 2989; DB 1; Length 3033;
```

```
Best Local Similarity 79.9%; Pred. No. 1.5e-197;
Matches 548; Conservative 63; Mismatches 75; Indels 0; Gaps 0;

Qy 1 MAPITAYAQOTRGLGCIITSLTGRDKNQVGEVQIVSTAAQTFLATCINGVCWTVVHGA 60
Db 1030 LAPITAYAQOTRGLGCIITSLTGRDKNQVGEVQIVSTAAQTFLATCINGVCWTVVHGA 1089
Qy 61 GTRTIASPKGPVIMQYTNVDLWGPAPQSGRSRLTPTCTCGSSDLXLVTRHADVIPVRR 120
Db 1090 GNKTLAGSRGPVTQMYSSAGDLVWGPSPPTKSLPCTCGAVDLVLTNRADVIPARR 1149
Qy 121 GDSRGSLLSRPPTSYLSKSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db 1150 GDXRGALLSRPRLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDIFIPVETLOIVT 1209
Qy 181 RSPVFTNSPPVPOQSFQVLAHLPATGSGKSTKVPAAQYKVLVLPNSVAATLGF 240
Db 1210 RSPVFTNSPPVPOQSFQVLAHLPATGSGKSTKVPAAQYKVLVLPNSVAATLGF 1269
Qy 241 AYMSKAHGIDPNIRTVGRTITGSPITYSTYKFLADGGCGGGAYDIIICDECHSDATS 300
Db 1270 AYLSKAHGINPNIRTVGRTITGAPITYSTYKFLADGGCAGGAYDIIICDECHAVDSTT 1329
Qy 301 ILGIGTVDQAEAGARLVLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db 1330 ILGIGTVDQAEAGARLVLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 1389
Qy 361 KGGRHLLFCHSKKCKDELAALVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYT 420
Db 1390 KGGRHLLFCHSKKCKDELAALRGMLNNAVAYRGLDVSVPITQGDVVVATDALMTGFT 1449
Qy 421 GDFSDVDCNTCVTQTVDVDFSLDPTFTTITLPODASVRSRRGRTGRGKPGIYRFVAPG 480
Db 1450 GDFSDVDCNVAVTQVDFSLDPTFTTITLPODASVRSRRGRTGRGLGIYRYVSTG 1509
Qy 481 ERPSGMFSDSVLCYDAGCAWYELTAEITTVRLRAYMNTPGLPVCDHLEFWEVGTGL 540
Db 1510 ERASGMFSDSVLCYDAGCAWYELTAEITTVRLRAYMNTPGLPVCDHLEFWEVGTGL 1569
Qy 541 THIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDQMWKCLIRLKLPTLHGPTPL 600
Db 1570 THIDAHFLSQTQKSGENFAYLTAYQVATVCARAKAPPPSDQMWKCLIRLKLPTLHGPTPL 1629
Qy 601 YRLGAVQNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
Db 1630 YRLGAVQNEITLTHPTVKYIATCQADLEVTSTWLVGGVLAALAAAYCLATGCVCIIGR 1689
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686
Db 1690 LHVNQRAVAVAPDKREVLYEAFDEMEEC 1715

RESULT 11
GNVJ08
genome polyprotein - hepatitis C virus (strain HC-38)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
C:Species: hepatitis C virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C:Accession: A40250; PQ0397; PQ0559
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.; J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to repository
A:Reference number: A40250; MUID:92230232; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0397
```

A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D10134  
R:Kato, N.; Ootsubayama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: PQ0554; MUID:92068204; PMID:1720309  
A:Accession: PQ0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BA001418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein C; status predicted <GPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <WEB>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis virus #status predicted <NS3>  
F:1234-1241/Product: nucleotide-binding motif A (P-loop)  
F:1316-1321/Product: nucleotide-binding motif B  
F:1320-1323/Product: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 82.3%; Score 2979; DB 1; Length 3033;  
Best Local Similarity 79.7%; Pred. No. 7.3e-177;  
Matches 547; Conservative 62; Mismatches 17; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTRGLGCIITSLTRDKNQVEGEVOIVSTAAOFLATCINGVMTVYHGA 60  
DB 1030 LAPITAYTOQTRGLGAIIVSLTRDKNEAQGVQLSSVYTOFLGTSIGVLWTVYHGA 1089

QY 61 GRTIASPKGPVQMTNVDDLVGWPAPQSSRLTCTCGSDLVLTVRHADVIPVRR 120  
DB 1090 GNTLAGPKGPVQMTVTSAGDLVGNPSPCTGSLDCTCGAVDLVLTNRADVIPVRR 1149

QY 121 GDSRGLSPRPISYLGKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1150 DRRGALLSPRLSTLGKSGGVPVLCRSHAGVGLFRAAVCARGVAKSIDRIPVESLDVAT 1209

QY 181 RSPVFTDNSPPVQSFQVAHLHAPGSGKSTKVPAAQGVKVLVLPNSVAATLGFG 240  
DB 1210 RSPFSFSDNSTPPVQSFQVGYLHAPGSGKSTKVPAAQGVKVLVLPNSVAATLGFG 1269

QY 241 AYMSKAHGDIPNRTGRTITTSPTITSTYTYGKFLADGGCGGAYDIIICDECHSTDATS 300  
DB 1270 AYMSKAHGINPRTGRTITTSPTITSTYTYGKFLADGGCGGAYDIIICDECHSTDATS 1329

QY 301 ILGIGTVLDOAETAGARLVVLTATPPGVSVPVPHNIEEVALSTTGIPYGAIPLEVI 360  
DB 1330 ILGIGTVLDOAETAGARLVVLTATPPGVSVPVPHNIEEVALSTTGIPYGAIPLEVI 1389

QY 361 KGRHLIFCHSKKCDLAALRGVNAVAYRGLDVSVPITGDDVVVATDALMTGYT 420  
DB 1390 KGRHLIFCHSKKCDLAALRGVNAVAYRGLDVSVPITGDDVVVATDALMTGYT 1449

QY 421 GDFSDVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRTORGRGTGRGKPGIYRFVAPG 480  
DB 1450 GDFSDVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRTORGRGTGRGKPGIYRFVAPG 1509

QY 481 ERPSGMFSDSVLCECYDAGCAWYELTTPAETTVLRAYNNTPGLPVCDHLEFEGVVTGL 540  
DB 1510 ERPSGMFSDSVLCECYDAGCAWYELTTPAETTVLRAYNNTPGLPVCDHLEFEGVVTGL 1569

QY 541 THIDAHFLSQTQKSGENLPYLVAQVATVCARAKAPPPSDOMWCKLIRKPLTHGPTPL 600  
DB 1570 THIDAHFLSQTQKSGENLPYLVAQVATVCARAKAPPPSDOMWCKLIRKPLTHGPTPL 1629

QY 601 YRLGAVQNEITLTHPVTKYIMTMSADLEVTSTWLVGGVLAALAAAYCLSTGCVCWIVGR 660

DB 1630 YRLGAVTNEVTLTHPVTKYIATCMQADLEIMTSSWVLVAGGVLAAYAAVCLATGTCISILGR 1689

QY 661 VVLSCKPAIIPREVLYREFDEMBEC 686  
DB 1690 LHLNDRVVVAPDKELIYEAFADEMBEC 1715

RESULT 12  
PC2219  
polyprotein - hepatitis C virus (type 5a) (fragments)  
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A protei  
C:Species: hepatitis C virus  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
A:Accession: PC2219  
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the he  
A:Reference number: PC2219; MUID:94338342; PMID:7520237  
A:Accession: PC2219  
A:Molecule type: mRNA  
A:Residues: 1-876 <STU>  
A:Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579  
A:Experimental source: serum  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: glycoprotein  
F:1-191/Product: core #status predicted <COE>  
F:68-78/Region: variable  
F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F:248-338/Region: E2  
F:339-411/Region: NS1 (amino end)  
F:412-783/Product: NS3 #status predicted <NSR>  
F:784-837/Product: NS4A #status predicted <NSA>  
F:838-876/Product: NS4B #status predicted <NSB>  
F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 55.1%; Score 1993; DB 2; Length 876;  
Best Local Similarity 85.5%; Pred. No. 1.6e-129;  
Matches 365; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

QY 260 ITTGSPITSTYGYKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDOAETAGARLV 319  
DB 411 ITTGASITSTYGYKFLADGGCGGAYDIIICDECHSQDATTILGIGTVLDOAETAGARLV 470

QY 320 VLATATPPGVSVPVPHNIEEVALSTTGIPYGAIPLEVIKGRHLIFCHSKKCDLA 379  
DB 471 VLATATPPGVSVPVPHNIEEVALPQGEVFPYGAIPLEVIKGRHLIFCHSKKCDLA 530

QY 380 AKLVALGINAVAYRGLDVSVPITGDDVVVATDALMTGYTGDPSVIDCNTCTVTQTVDF 439  
DB 531 KQLTSLGVNAVAYRGLDVAIPTAGDGVVVCSTDALMTGTFGDSVIDCNSAVTQTVDF 590

QY 440 SLDPFTTETITLPQDAVSRTORGRGTGRGKPGIYRFVAPGERSGMFSDSVLCECYDAG 499  
DB 591 SLDPFTTETITLPQDAVSRSQRGRGTGRGKPGIYRFVAPGERSGMFSDSVLCECYDAG 650

QY 500 CANYELTTPAETTVLRAYNNTPGLPVCDHLEFEGVFTGLTHIDAHFLSQTQKSGENLP 559  
DB 651 CANYDLTTPAETTVLRAYNNTPGLPVCDHLEFEGVFTGLTHIDAHFLSQTQKSGENFP 710

QY 560 YLVAQVATVCARAKAPPPSDOMWCKLIRKPLTHGPTPLLYRLGAVQNEITLTHPVTKY 619  
DB 711 YLVAQVATVCARAKAPPPSDMTWCKLIRKPLTGTPTPLLYRLGAVQNEITLTHPVTKY 770

QY 620 IMTMSADLEVTSTWLVGGVLAALAAAYCLSTGCVCWIVGRVLSGKPAIIPREVLYRE 679  
DB 771 IMACMSADLEVTSTWLVGGVVAALAAAYCLTGVSAIVGRILISGKPAIIPREVLYQQ 830

QY 680 FDEMBEC 686  
DB 831 FDEMBEC 837

RESULT 13  
S68016  
ATPase/RNA helicase - hepatitis C virus (fragment)  
C;Species: hepatitis C virus  
C;Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: S68016  
R;Jin, L.; Peterson, D.L.  
Arch. Biochem. Biophys. 323, 47-53, 1995  
A;Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
A;Reference number: S68016; MUID:96019946; PMID:7487072  
A;Accession: S68016  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-386 <JIN>  
A;Cross-references: UNIPROT:Q04045  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polypeptide  
F;24-31/Region: nucleotide-binding motif A (P-loop)  
F;86-91/Region: nucleotide-binding motif B  
F;90-93/Region: DEXH motif

Query Match 54.5%; Score 1972.5; DB 2; Length 386;  
Best Local Similarity 92.4%; Pred. No. 1.4e-128;  
Matches 375; Conservative 3; Mismatches 7; Indels 21; Gaps 2;

Qy 183 PVFTDN--SSPPVPPQSFQVAHLHAPTGGKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 241  
Db 1 PVFTDNSSPPVPPQSFQVAHLHAPTGGKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 60  
Qy 242 YMSKAHGIDNIRGVRTITTTGSPITVSTYKGLADGGCGGAYDIIICDECHSTDATSI 301  
Db 61 YMSKAHGVD-----YCKFLADGGCGGAYDIIICDECHSTDATSI 100  
Qy 302 LGTGTVLDOAETAGARLVATATPGSVTVPHNIEEVALSTTGEIPFYKALPLEVIK 361  
Db 101 LGTGTVLDOAETAGARLVATATPGSVTVPHNIEEVALSTTGEIPFYKALPLESAIK 160  
Qy 362 GGRHLIFCHSKKKCKDELAALVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 421  
Db 161 GGRHLIFCHSKKKCKDELAALVALGINAVAYRGLDVSVIPSSGDVVVATDALTMTGFSG 220  
Qy 422 DFDSVIDCNTCVTQVDFSLDPTTETITLPDAVSRRTORRGTRGKPGIYRFVAPGE 481  
Db 221 DFDSVIDCNTCVTQVDFSLDPTTETITLPDAVSRRTORRGTRGKPGIYRFVAPGE 280  
Qy 482 RPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLT 541  
Db 281 RPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLT 340  
Qy 542 HIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLI 587  
Db 341 HIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLI 386

RESULT 14  
PS0326  
polypeptide - hepatitis C virus (isolate Fla) (fragments)  
C;Species: hepatitis C virus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: PS0326  
R;Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.  
Gene 105, 167-172, 1991  
A;Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype  
A;Reference number: PS0326; MUID:92039028; PMID:1718820  
A;Accession: PS0326  
A;Molecule type: Genomic RNA  
A;Residues: 1-492 <LIJ>  
A;Cross-references: UNIPROT:Q91PES; UNIPROT:Q36579; UNIPROT:Q036610; UNIPROT:Q03463; UNIPROT:Q0320  
A;Note: this sequence corresponds to nonstructural protein NS3 region  
A;Note: translation of the nucleotide sequence is not complete  
C;Superfamily: hepatitis C virus genome polypeptide

C;Keywords: polypeptide

Query Match 38.4%; Score 1388; DB 2; Length 492;  
Best Local Similarity 98.1%; Pred. No. 4.1e-88;  
Matches 257; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 425 SVIDCNTCVTQVDFSLDPTTETITLPDAVSRRTORRGTRGKPGIYRFVAPGERPS 484  
Db 1 SVIDCNTCVTQVDFSLDPTTETITLPDAVSRRTORRGTRGKPGIYRFVAPGERPS 60  
Qy 485 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLTHID 544  
Db 61 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLTHID 120  
Qy 545 AHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLIIRLKLTHGPTPLLYRLG 604  
Db 121 AHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLIIRLKLTHGPTPLLYRLG 180  
Qy 605 AVQNEITLTHPVTKYIMTCHSADLEVVSTWLVGGVLAALAAAYCLSTGCWVIVGRVLS 664  
Db 181 AVQNEITLTHPVTKYIMTCHSADLEVVSTWLVGGVLAALAAAYCLSTGCWVIVGRVLS 240  
Qy 665 GKPAIIPDREVLYREFDEMEEC 686  
Db 241 GKPAIIPDREVLYREFDEMEEC 262

RESULT 15

JQ1366

polypeptide - hepatitis C virus (French isolate) (fragments)

C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C;Accession: JQ1366

R;Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A;Reference number: JQ1366; MUID:92013977; PMID:1655961

A;Accession: JQ1366

A;Molecule type: Genomic RNA

A;Residues: 1-716 &lt;KRE&gt;

A;Cross-references: UNIPROT:Q9PX22

C;Superfamily: hepatitis C virus genome polypeptide

C;Keywords: Glycoprotein; polypeptide

F;84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 38.2%; Score 1383; DB 2; Length 716;

Best Local Similarity 51.8%; Pred. No. 1.5e-87;

Matches 318; Conservative 43; Mismatches 103; Indels 150; Gaps 21;

Qy 132 PISYKSGSGGLLCPAGHAGIFRAAVCTRGVAKAVDFIPVENLETMRSPVFTDNSSP 191

Db 138 PISYANGT--GPEHRP-----YCHWYPPKPGIYPAQ---TVCGPVYCFPTSP 180

Qy 192 PVYPSQFQVAHLHAPTGGKSTKVPAAAYAAQGYKVLNPSVAATLGL--FG-----AYM 243

Db 181 VVVGTTNK---LGAPTYN-----WGNDDTDVFLN--NTRPPLGNWFCCTWVNSGF 227

Qy 244 SKAHGIDPNIRTVRTITTTGSP-----TSTYTGKFLADGGCGGAYDIIICDECHSTD 297

Db 228 TKVCGAPCVIGGAGNNTLYCPDCKPKHPEATYSR-----CGSGPW---ITPRC----- 274

Qy 298 ATSLG-----IGTV-----LDQATAGARL 318

Db 275 ---LVGYRPLMHPCTVNTLTKVRYMVGVEHRLQVACNWTGRGNCNDRDRSELSP 331

Qy 319 VLAT-----ATPGSVTVPHNIEEVALSTTGEIPFYKALPLEVIKGRHLIFCHSKKK 374

Db 332 LLSTQWQVLPSCFTTLP-----AL--TGLHLHQNIVDVQVLYG----- 371

Qy 375 CDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYTGFDSVIDC--NTC 432

Db 372 -----VGSGIVSWAKWEVII-----LFLLLADA-----RVCCLWNTC 406

QY	433	VTQTVDFSLDPTFTIETITLPQDAVSRTORRGRTGRGKPGIYRFVAPGERPSCGMFDSSVL	492
Db	407	VTQTVDFSLDPTFTIETITLPQDAVSRTORRGRTGRGKPGIYRFVAPGERPSCGMFDSSVL	466
QY	493	CECYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSQT	552
Db	467	CECYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSQT	526
QY	553	QSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKP TLHGPTPLLYRLGAVQNEITL	612
Db	527	QSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKP TLHGPTPLLYRLGAVQNEVIL	586
QY	613	THPVTKYIMTCSADLEWVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRVVLSCGP	672
Db	587	THPITKYIMTCSADLEWVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRVVLSCGP	646
QY	673	REVLRYREFDEMEEC	686
Db	647	REVLRYREFDEMEEC	660

Search completed: November 7, 2005, 20:10:53  
Job time : 17.1714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 87.9151 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-2

Perfect score: 3619

Sequence: 1 MAPITAYAAQTRGLGLCIIT.....PAIIPREVLYREFDEMEEC 686

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3602	99.5	2436	2 Q81756	Q81756 hepatitis c
2	3602	99.5	3011	1 POLG_HCV1	P26664 h genome po
3	3598	99.4	3011	2 Q1FPE5	Q1FPE5 hepatitis c
4	3574	98.8	2908	2 Q61X04	Q61X04 hepatitis c
5	3560	98.4	3011	2 Q35579	Q35579 hepatitis c
6	3553	98.2	3011	2 Q36608	Q36608 hepatitis c
7	3553	98.2	3015	2 Q9PMU9	Q9PMU9 hepatitis c
8	3553	98.2	3015	2 Q9PMU5	Q9PMU5 hepatitis c
9	3549	98.1	3011	2 Q9EL58	Q9EL58 hepatitis c
10	3542	97.9	3011	2 Q36609	Q36609 hepatitis c
11	3542	97.9	3011	2 Q36610	Q36610 hepatitis c
12	3542	97.9	3011	2 Q03463	Q03463 hepatitis c
13	3540	97.8	3011	2 Q9DIT6	Q9DIT6 hepatitis c
14	3513	97.1	3011	1 POLG_HCVH	P27958 h genome po
15	3436	94.9	3010	2 Q9DTE2	Q9DTE2 hepatitis c
16	3436	94.9	3010	2 Q9QIY2	Q9QIY2 hepatitis c
17	3433	94.9	3010	2 Q9QIY1	Q9QIY1 hepatitis c
18	3432	94.8	3010	2 Q9J3G2	Q9J3G2 hepatitis c
19	3426	94.7	3010	1 POLG_HCVBK	P26663 h genome po
20	3426	94.7	3010	2 Q9DTE7	Q9DTE7 hepatitis c
21	3426	94.7	3010	2 Q9QIX2	Q9QIX2 hepatitis c
22	3425	94.6	3010	2 Q9SAU2	Q9SAU2 hepatitis c
23	3425	94.6	3010	2 Q68788	Q68788 hepatitis c
24	3425	94.6	3010	2 Q9J3H6	Q9J3H6 hepatitis c
25	3425	94.6	3013	2 Q636P5	Q636P5 hepatitis c
26	3424	94.6	3010	2 Q9J3H0	Q9J3H0 hepatitis c
27	3424	94.6	3010	2 Q9J3H8	Q9J3H8 hepatitis c
28	3422	94.6	3013	2 Q9QNC0	Q9QNC0 hepatitis c
29	3421	94.5	3010	2 Q9QIX1	Q9QIX1 hepatitis c
30	3421	94.5	3010	2 Q9QIY9	Q9QIY9 hepatitis c
31	3421	94.5	3010	2 Q9QIZ0	Q9QIZ0 hepatitis c

#### RESULT 1

ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; asRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus			
OX	NCBI_taxid=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: M32084; AAA45677.1; -			
DR	PIR: PS0326; PS0326.			
DR	PIR: PS0327; PS0328.			
DR	PIR: PS0328; PS0328.			
DR	HSSP: P27958; 1A1V.			
DR	GO: GO:0016021; C:integral to membrane; IEA.			
DR	GO: GO:0019038; C:viral capsid; IEA.			
DR	GO: GO:0019031; C:viral envelope; IEA.			
DR	GO: GO:0005524; F:ATP binding; IEA.			
DR	GO: GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO: GO:0003723; F:RNA binding; IEA.			
DR	GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO: GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO: GO:0005199; F:structural molecule activity; IEA.			
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO: GO:0006350; P:transcription; IEA.			
DR	GO: GO:0019079; P:viral genome replication; IEA.			
DR	GO: GO:0019087; P:viral transformation; IEA.			
DR	InterPro: IPR000345; CytC_heme_BS.			
DR	InterPro: IPR001410; DEAD.			
DR	InterPro: IPR011545; DEAD/DEAH_N.			
DR	InterPro: IPR002531; HCV NS1.			
DR	InterPro: IPR000745; HCV NS4a.			
DR	InterPro: IPR001490; HCV NS4b.			
DR	InterPro: IPR002868; HCV NS5a.			
DR	InterPro: IPR002166; HCV NS5b.			
DR	InterPro: IPR001650; Helicase_C.			
DR	InterPro: IPR004109; Peptidase_S29.			
DR	InterPro: IPR009003; Pept_Ser_Cys.			
DR	InterPro: IPR002518; Pept_U39_HCV_NS2.			
DR	InterPro: IPR007094; RNA_pol_Fsvir.			
DR	Pfam: PF01560; HCV NS1; 1.			
DR	Pfam: PF01538; HCV NS2; 1.			
DR	Pfam: PF02907; HCV NS3; 1.			
DR	Pfam: PF01006; HCV NS4a; 1.			
DR	Pfam: PF01001; HCV NS4b; 1.			
DR	Pfam: PF01506; HCV NS5a; 1.			

#### ALIGNMENTS

32	3421	94.5	3013	2	Q9QIY0	Q9QIY0 hepatitis c
33	3421	94.5	3014	2	Q6GYR8	Q6GYR8 hepatitis c
34	3420	94.5	3010	2	Q9DTE5	Q9DTE5 hepatitis c
35	3420	94.5	3010	2	Q9J3G8	Q9J3G8 hepatitis c
36	3420	94.5	3010	2	Q9J3G9	Q9J3G9 hepatitis c
37	3420	94.5	3013	2	Q9QIX9	Q9QIX9 hepatitis c
38	3419	94.5	3010	2	Q6GYR9	Q6GYR9 hepatitis c
39	3419	94.5	3010	2	Q9DTE0	Q9DTE0 hepatitis c
40	3419	94.5	3010	2	Q9QIY5	Q9QIY5 hepatitis c
41	3419	94.5	3010	2	Q9QIY6	Q9QIY6 hepatitis c
42	3419	94.5	3015	2	Q9WPH5	Q9WPH5 hepatitis c
43	3418	94.4	3010	2	Q9DTE6	Q9DTE6 hepatitis c
44	3418	94.4	3010	2	Q9J3F9	Q9J3F9 hepatitis c
45	3417	94.4	3010	1	POLG_HCVJT	Q00269 h genome po





KW Envelope protein; Glycoprotein; Helicase; Hydrolase;  
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
KW Serine protease; Transferase; Transmembrane.  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3011  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 476 476  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2364 2364  
FT CARBOHYD 2789 2789  
FT STRAND 1224 1226  
FT TURN 1232 1234  
FT TURN 1236 1238  
FT HELIX 1239 1246  
FT TURN 1247 1248  
FT STRAND 1251 1255  
FT STRAND 1291 1295  
FT HELIX 1296 1301  
FT TURN 1302 1303  
FT TURN 1312 1315  
FT TURN 1316 1319  
FT HELIX 1323 1335  
FT TURN 1336 1336  
FT HELIX 1337 1339  
FT TURN 1340 1340  
FT STRAND 1343 1346  
FT TURN 1352 1353  
FT STRAND 1363 1366  
FT STRAND 1373 1373  
FT TURN 1376 1377  
FT STRAND 1380 1380  
FT STRAND 1382 1385  
FT STRAND 1390 1393  
FT HELIX 1397 1408  
FT TURN 1409 1411  
FT STRAND 1414 1417  
FT TURN 1430 1431  
FT STRAND 1433 1436  
FT TURN 1438 1439  
FT TURN 1442 1444  
FT STRAND 1450 1453  
FT STRAND 1456 1463  
FT STRAND 1471 1478

FT HELIX 1481 1488  
FT TURN 1489 1490  
FT STRAND 1498 1501  
FT HELIX 1514 1526  
FT TURN 1527 1527  
FT HELIX 1532 1544  
FT TURN 1546 1547  
FT HELIX 1553 1563  
FT TURN 1564 1564  
FT HELIX 1570 1578  
FT TURN 1579 1580  
FT HELIX 1584 1597  
FT TURN 1598 1598  
FT HELIX 1606 1611  
FT TURN 1614 1618  
FT STRAND 1622 1623  
FT STRAND 1627 1627  
FT STRAND 1635 1636  
FT HELIX 1640 1652  
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;  
Query Match 99.5%; Score 3602; DB 1; Length 3011;  
Best Local Similarity 99.6%; Pred. No. 8.3e-244;  
Matches 583; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MAPTAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60  
DB 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 1085  
QY 61 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120  
DB 1086 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145  
QY 121 GDSRGSLLSPRPISYLVKSGSGPLLCPPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1146 GDSRGSLLSPRPISYLVKSGSGPLLCPPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
QY 181 RSPVFTDNSSPPVPPQSPQVAHLHAPTQSGSKSTKVPAAYAAQGYKVLVLPNSVAATLGF 240  
DB 1206 RSPVFTDNSSPPVPPQSPQVAHLHAPTQSGSKSTKVPAAYAAQGYKVLVLPNSVAATLGF 1265  
QY 241 AYMSKAHGIDNIRTVGRTITTTGSPITYTSTYKFLADGCGSGGAYDIIICDECHSTDATS 300  
DB 1266 AYMSKAHGIDNIRTVGRTITTTGSPITYTSTYKFLADGCGSGGAYDIIICDECHSTDATS 1325  
QY 301 ILGIGTVLDQAEATAGARLVVLATATPPGSGVTPVPHNIEEVALSTTGEIPFYGKAIPLEVI 360  
DB 1326 ILGIGTVLDQAEATAGARLVVLATATPPGSGVTPVPHNIEEVALSTTGEIPFYGKAIPLEVI 1385  
QY 361 KGGRHILFCHSKKKCDELAALKLVALGINAVAYYRGLDVSVIPPIQDVVVVATDALMTGYT 420  
DB 1386 KGGRHILFCHSKKKCDELAALKLVALGINAVAYYRGLDVSVIPPIQDVVVVATDALMTGYT 1445  
QY 421 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTORRGTRGKPGIYRFVAPG 480  
DB 1446 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTORRGTRGKPGIYRFVAPG 1505  
QY 481 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPYCQDHFLEWEGVFTGL 540  
DB 1506 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPYCQDHFLEWEGVFTGL 1565  
QY 541 THIDAHFLSQKSGENIPYLVAQATVCARAQAPPPSWDQWKKLIRLKPTLHGPTPL 600  
DB 1566 THIDAHFLSQKSGENIPYLVAQATVCARAQAPPPSWDQWKKLIRLKPTLHGPTPL 1625  
QY 601 YRLGAVQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSGCVVIVGR 660  
DB 1626 YRLGAVQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSGCVVIVGR 1685  
QY 661 VVLSGKPAIIPDREVLXREFDEMEEC 686  
DB 1686 VVLSGKPAIIPDREVLXREFDEMEEC 1711

```
RESULT 3
Q91FES PRELIMINARY; PRT; 3011 AA.
AC Q91FES;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
DR EMBL: AF271632; AAF81759.1; -.
DR PIR: A4150; A44150.
DR PIR: Q0804; Q0804.
DR PIR: PS0326; PS0326.
DR PIR: PS0327; PS0327.
DR PIR: PS0328; PS0328.
DR HSP: O8JYS1; 1CWX.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003668; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR00345; CytC_heme_BS.
DR InterPro: IPR01410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH N.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_S29.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR InterPro: IPR002518; Pept_U39_HCV_NS2.
DR InterPro: IPR007095; RNA_pol_Ds_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;
```

```
Query Match 99.4%; Score 3598; DB 2; Length 3011;
Best Local Similarity 99.4%; Pred. No. 1.6e-243; Indels 0; Gaps 0;
Matches 682; Conservative 2; Mismatches 2;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATFCINGVCMWTVYHGA 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATFCINGVCMWTVYHGA 1085

Qy 61 GTRTIASPKGPVIOQMTNVDDLVGHPAPOGSSSLTPTCTGSSDLYLVTSHADVIPVRRR 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 GTRTIASPKGPVIOQMTNVDDLVGHPAPOGSSSLTPTCTGSSDLYLVTSHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRISYLLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1146 GDSRGSLLSPRISYLLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPSPVAATLGFG 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1206 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPSPVAATLGFG 1265

Qy 241 AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1266 AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 1325

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGVTVPHPNIEEVALSTTGIPFVYKAIPLLEVI 360
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1326 ILGIGTVLDOAETAGARLVVLTATPPGVTVPHPNIEEVALSTTGIPFVYKAIPLLEVI 1385

Qy 361 KGRHLIFCHSKKKCKDELAALVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 420
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1386 KGRHLIFCHSKKKCKDELAALVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 1445

Qy 421 GDFDSVIDCNTCTVTVDFSLDFTFETITLPODAVSRQRTGRGRTGKGGYRFPVAPG 480
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1446 GDFDSVIDCNTCTVTVDFSLDFTFETITLPODAVSRQRTGRGRTGKGGYRFPVAPG 1505

Qy 481 ERPSGMFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 540
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1506 ERPSGMFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 1565

Qy 541 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMMKCLIRLKTLLHGPTPLL 600
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1566 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMMKCLIRLKTLLHGPTPLL 1625

Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWLVGGVLAALAAAYCLSTGCVNLVGR 660
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWLVGGVLAALAAAYCLSTGCVNLVGR 1685

Qy 661 VLSGKPAIIPDREVLYRPEDEMEEC 686
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1686 VLSGKPAIIPDREVLYRPEDEMEEC 1711

RESULT 4
Q6IX04 PRELIMINARY; PRT; 2908 AA.
AC Q6IX04;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1a;
RA Brann T.W., Korttilä S., Polis M., Imamichi T.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY615798; AAT44836.1; -.
DR HSSP: P27958; 1A1V.
```

DR GO: 0016021; C: integral to membrane; IEA.  
DR GO: 0019028; C: viral capsid; IEA.  
DR GO: 0019031; C: viral envelope; IEA.  
DR GO: 0005524; F: ATP binding; IEA.  
DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
DR GO: 0016787; F: hydrolase activity; IEA.  
DR GO: 0003723; F: RNA binding; IEA.  
DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO: 0008236; F: serine-type peptidase activity; IEA.  
DR GO: 0005198; F: structural molecule activity; IEA.  
DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
DR GO: 0006350; P: transcription; IEA.  
DR GO: 0019079; P: viral genome replication; IEA.  
DR GO: 0019087; P: viral transformation; IEA.  
DR InterPro: IPR00345; CytC\_heme\_BS.  
DR InterPro: IPR011410; DEAD.  
DR InterPro: IPR011545; DEAD/DEAH\_N.  
DR InterPro: IPR02522; HCV\_capsid.  
DR InterPro: IPR02521; HCV\_core.  
DR InterPro: IPR02519; HCV\_env.  
DR InterPro: IPR02531; HCV\_NS1.  
DR InterPro: IPR00745; HCV\_NS1a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_NS5a.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; Helicase\_C; 1.  
DR Pfam: PF00998; Viral RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.  
FT NON\_TER 2308 2308  
SQ SEQUENCE 2308 AA; 315737 MW; BFS44BC591498A4F CRC64;  
Query Match 98.8%; Score 3574; DB 2; Length 2908;  
Best Local Similarity 97.7%; Pred. No. 7.4e-242;  
Matches 670; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MAPITAAQTRGLGCIITSLTRDKNQEVEQIVSTAAQTFLATCINGCVTVYHGA 60  
Db 1026 LAPITAAQTRGLGCIITSLTRDKNQEVEQIVSTAAQTFLATCINGCVTVYHGA 1085  
Qy 61 GTTIIASPKGVIMYNTVDQDLGWPAQPSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GTTIIASPKGVIMYNTVDQDLGWPAQPSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145  
Qy 121 GDSRGLSPRPISYLVKSGSGGLPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGLSPRPISYLVKSGSGGLPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVQSFQVAHLHAPTSKGSTKVPAAQGVKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVQSFQVAHLHAPTSKGSTKVPAAQGVKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGIDNIRTVGRTITGSPITYSTYKFLADGGCGGAYDIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDNIRTVGRTITGSPITYSTYKFLADGGCGGAYDIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDQAETAGARLVLAATATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEVI 360

Db 1326 ILGIGTVLDQAETAGARLVLAATATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEAI 1385  
Qy 361 KGGRLIFCHSKKKCDLAALVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 420  
Db 1386 KGGRLIFCHSKKKCDLAALVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGFT 1445  
Qy 421 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLQDAVSRQRTGRCGKGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLQDAVSRQRTGRCGKGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCCEVDAGCAWYELPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 540  
Db 1506 ERPSGMFDSVLCCEVDAGCAWYELPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 1565  
Qy 541 THIDAHFLSQTKSGENLPYLVAIQATVCARAQPPSDQMCKLIRLXPTLHGPTPL 600  
Db 1566 THIDAHFLSQTKSGENLPYLVAIQATVCARAQPPSDQMCKLIRLXPTLHGPTPL 1625  
Qy 601 YRLGAVONEITLTHPVTKYIMTCKSADLEVTSTWLVGGVLAALAAAYCLSTGCWIVGR 660  
Db 1626 YRLGAVONEITLTHPVTKYIMTCKSADLEVTSTWLVGGVLAALAAAYCLSTGCWIVGR 1685  
Qy 661 VVLSGKPAIIPDREVLYRBEFDEMEEC 686  
Db 1686 IVLSGKPAVIPDREVLYRBEFDEMEEC 1711  
RESULT 5  
O36579 PRELIMINARY; PRT; 3011 AA.  
AC O36579;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
CX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;  
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,  
Rice C.M.;  
RT "Transmission of hepatitis C by intrahepatic inoculation with  
transcribed RNA.";  
RL Science 277:570-574(1997).  
DR EMBL; AF009606; AB66324.1; -.  
DR PIR; A44150; A44150.  
DR PIR; PQ0804; PQ0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR PDB; INLL; X-ray; A/B=1017-1214.  
DR GO: 0019028; C: viral capsid; IEA.  
DR GO: 0019031; C: viral envelope; IEA.  
DR GO: 0005524; F: ATP binding; IEA.  
DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
DR GO: 0003723; F: RNA binding; IEA.  
DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO: 0008236; F: serine-type peptidase activity; IEA.  
DR GO: 0005198; F: structural molecule activity; IEA.  
DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
DR GO: 0006350; P: transcription; IEA.  
DR GO: 0019079; P: viral genome replication; IEA.  
DR GO: 0019087; P: viral transformation; IEA.  
DR InterPro; IPR00345; CytC\_heme\_BS.  
DR InterPro; IPR011410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR02522; HCV\_capsid.  
DR InterPro; IPR02521; HCV\_core.  
DR InterPro; IPR02519; HCV\_env.

DR	InterPro; IPR002531; HCV_NSI1.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR001490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRP.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR004109; Peptidase_S29.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF01543; HCV_capsid; 1.	
DR	Pfam; PF01542; HCV_core; 1.	
DR	Pfam; PF01539; HCV_env; 1.	
DR	Pfam; PF01538; HCV_NS1; 1.	
DR	Pfam; PF01538; HCV_NS2; 1.	
DR	Pfam; PF01538; HCV_NS3; 1.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF01006; HCV_NS4a; 1.	
DR	Pfam; PF01506; HCV_NS4b; 1.	
DR	Pfam; PF01001; HCV_NS4b; 1.	
DR	Pfam; PF00271; Helicase_C; 1.	
DR	Pfam; PF00998; Viral_RdRP; 1.	
DR	SMART; SM00487; DEXDC; 1.	
DR	PROSITE; PS00140; CYTOCHROME C; UNKNOWN 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; Transmembrane.	
SQ	SEQUENCE 3011 AA; 327184 MW; E2R0EB809C63C1B9 CRC64;	
Query Match 98.4%; Score 3560; DB 2: Length 3011;		
Best Local Similarity 97.7%; Pred. No. 7.5e-241;		
Matches 670; Conservative 9; Mismatches 7; Indels 0; Gaps 0;		
QY	1 MAPITAYAOQTRGLGCIITSLTGRDNQVGEVQIVSTAAQTFLATPINGVCTWYVHGA 60	
DB	1026 LAPITAYAOQTRGLGCIITSLTGRDNQVGEVQIVSTATQIFLATPINGVCTWYVHGA 1085	
QY	61 GTTITASPKGPVIMQYTNVDLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120	
DB	1086 GTRTITASPKGPVIMQYTNVDLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145	
QY	121 GDSRGSLLSPRPISYLGKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180	
DB	1146 GDSRGSLLSPRPISYLGKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205	
QY	181 RSPVFTDSSPPVPSQFVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLGF 240	
DB	1206 RSPVFTDSSPPVPSQFVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLGF 1265	
QY	241 AYMSKAHGDIPNIRTGVRTTTTGSPTITYSTYGFKLADGGCGGGAYDIIICDECHSTDATS 300	
DB	1266 AYMSKAHGDIPNIRTGVRTTTTGSPTITYSTYGFKLADGGCGGGAYDIIICDECHSTDATS 1325	
QY	301 ILGIGTVLDQAEATAGARLVLATATPGSVTVRHPNIEEVALSTTGEIPFGKAIPL 360	
DB	1326 ILGIGTVLDQAEATAGARLVLATATPGSVTVRHPNIEEVALSTTGEIPFGKAIPL 1385	
QY	361 KGGRHLLIFCHSKKCDLAALVALGINAVAYRGLDVSVIPTGDDVVVATDALMTGYT 420	
DB	1386 KGGRHLLIFCHSKKCDLAALVALGINAVAYRGLDVSVIPTGDDVVVATDALMTGYT 1445	
QY	421 GDFSDVTDNCTVQTQVDFSLDPTFTTETTLPODAVSRQRRGRTGRGPIYRFVAPG 480	
DB	1446 GDFSDVTDNCTVQTQVDFSLDPTFTTETTLPODAVSRQRRGRTGRGPIYRFVAPG 1505	
QY	481 ERPSGMFSSVLCBCYDAGCAWYELTAEITTVRLRAYMNTPLGVCDHLEFEGVFTGL 540	
DB	1506 ERPSGMFSSVLCBCYDAGCAWYELTAEITTVRLRAYMNTPLGVCDHLEFEGVFTGL 1565	
QY	541 THIDAHFLSQTQSGENLPYLVAQYATVCARAQAPPSPSDQMMKCLIRLKPFLHGP 600	
DB	1566 THIDAHFLSQTQSGENFPYLVAQYATVCARAQAPPSPSDQMMKCLIRLKPFLHGP 1625	
QY	601 YRLGAVQNEITLTHPTVKYINTCMGADLEVTSTWLVGGVLAALAAAYCLSTGC	

DB	1626 YRLGAVQNEITLTHPTVKYINTCMGADLEVTSTWLVGGVLAALAAAYCLSTGC	1685
QY	661 VVLSGKPAIIPDREVLYREFDEMEEC 686	
DB	1686 IVLSGKPAIIPDREVLYQEFDEMEEC 1711	
RESULT 6		
O36608 PRELIMINARY; PRT; 3011 AA.		
ID	O36608	
AC	O36608;	
DT	01-JAN-1998 (TrEMBLrel. 05, Created)	
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Polyprotein.	
OS	Hepatitis C virus strain H77.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.	
OX	NCBI_TaxID=63746;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H77;	
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;	
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;	
RT	"Transcript from a single full-length cDNA clone of hepatitis C virus	
RT	are infectious when directly transfected into the liver of a	
RT	chimpanzee.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).	
DR	EMBL; AF011751; AAB67036.1; -.	
DR	PIR; A4150; A4150.	
DR	PIR; PQ0804; PQ0804.	
DR	PIR; PS0326; PS0326.	
DR	PIR; PS0327; PS0327.	
DR	PIR; PS0328; PS0328.	
DR	HSSP; P27958; 1HEI.	
DR	GO; GO:0019028; C:viral capsid; IEA.	
DR	GO; GO:0019031; C:viral envelope; IEA.	
DR	GO; GO:0005524; F:ATP binding; IEA.	
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.	
DR	GO; GO:0003723; F:RNA binding; IEA.	
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.	
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.	
DR	GO; GO:0005198; F:structural molecule activity; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	GO; GO:0006350; P:transcription; IEA.	
DR	GO; GO:0019079; P:viral genome replication; IEA.	
DR	GO; GO:0019087; P:viral transformation; IEA.	
DR	InterPro; IPR000345; CytC_heme_BS.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR011545; DEAD/DEAH_N.	
DR	InterPro; IPR002522; HCV_capsid.	
DR	InterPro; IPR002521; HCV_core.	
DR	InterPro; IPR002519; HCV_env.	
DR	InterPro; IPR002531; HCV_NS1.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR001490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRP.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR004109; Peptidase_S29.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF01543; HCV_capsid; 1.	
DR	Pfam; PF01542; HCV_core; 1.	
DR	Pfam; PF01539; HCV_env; 1.	
DR	Pfam; PF01538; HCV_NS1; 1.	
DR	Pfam; PF01538; HCV_NS2; 1.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF01006; HCV_NS4a; 1.	
DR	Pfam; PF01001; HCV_NS4b; 1.	







Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDDVVVWATDALMTGYT 420  
Db 1390 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPTSDDVVVWSTDALMTGFT 1449  
Qy 421 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRTOBGRGRTGKPGIYRFFVAPG 480  
Db 1450 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRTOBGRGRTGKPGIYRFFVAPG 1509  
Qy 481 ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEFEGVFTGL 540  
Db 1510 ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEFEGVFTGL 1569  
Qy 541 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPPSDQMKLRLKPTLHGPTPL 600  
Db 1570 THIDAHFLSQKSGENFPYLVAQVATVCARAQAPPPSDQMKLRLKPTLHGPTPL 1629  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCHSADLEVVTSWVLVGGVLAALAYCLSTGCWIVGR 660  
Db 1630 YRLGAVQNEITLTHPVTKYIMTCHSADLEVVTSWVLVGGVLAALAYCLSTGCWIVGR 1689  
Qy 661 VLSGKPAIIPDREVLRFDEMEEC 686  
Db 1690 VLSGKPAIIPDREVLRFDEMEEC 1715  
RESULT 9  
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.  
AC Q9ELS8  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polypeptide.  
OS Hepatitis C virus.  
OC Virusae; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Desai S.M., Devare S., Yamaguchi J.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF290978; AAG02099.1; -.  
DR PIR; A44150; A44150.  
DR PIR; PQ0804; PQ0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P26664; lHEI.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD\_N.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV env.  
DR InterPro; IPR002519; HCV NS1.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.

DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327108 MW; A6BECF5A3B3EE13F CRC64;  
Query Match 98.1%; Score 3549; DB 2: Length 3011;  
Best Local Similarity 97.5%; Pred. No. 4.4e-240;  
Matches 669; Conservative 9; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 MAPITAYAAQQTGRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db 1026 LAPITAYAAQQTGRGLGCIITSLTGRDKNQVEGEVQIVSTATQIFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPIVQMYTNVDQDLVGMWPAQGSRSRLTPTCGSSDLYLVTRHADVIPVRR 120  
Db 1086 GTRTIASPKGPIVQMYTNVDQDLVGMWPAQGSRSRLTPTCGSSDLYLVTRHADVIPVRR 1145  
Qy 121 GDSRGSLLSPRISYLSKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDPTPVENLETTM 180  
Db 1146 GDSRGSLLSPRISYLSKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDPTPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVVPQSFQVAHLHAPGTSGSKSTKVPAAAYAAQGYKVLVLNPSVAATLGFG 240  
Db 1206 RSPVFTDNSSPPVVPQSFQVAHLHAPGTSGSKSTKVPAAAYAAQGYKVLVLNPSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNIRTGVRTITTTGSPITYSYGKFLADGGCGSGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTGVRTITTTGSPITYSYGKFLADGGCGSGAYDIIICDECHSTDATS 1325  
Qy 301 ILGITGVLDQAGTAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360  
Db 1326 ILGITGVLDQAGTAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDDVVVWATDALMTGYT 420  
Db 1386 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPASGDDVVVWSTDALMTGFT 1445  
Qy 421 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRTOBGRGRTGKPGIYRFFVAPG 480  
Db 1446 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRTOBGRGRTGKPGIYRFFVAPG 1505  
Qy 481 ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEFEGVFTGL 540  
Db 1506 ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEFEGVFTGL 1565  
Qy 541 THIDAHFLSQKSGENLPYLVAQVATVCARAQAPPPSDQMKLRLKPTLHGPTPL 600  
Db 1566 THIDAHFLSQKSGENFPYLVAQVATVCARAQAPPPSDQMKLRLKPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCHSADLEVVTSWVLVGGVLAALAYCLSTGCWIVGR 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCHSADLEVVTSWVLVGGVLAALAYCLSTGCWIVGR 1685  
Qy 661 VLSGKPAIIPDREVLRFDEMEEC 686  
Db 1690 VLSGKPAIIPDREVLRFDEMEEC 1715

DB	1686	IVLSGKPAIIPDREVLYQSFDEMEEC	1711	
RESULT 10				
O36609		PRELIMINARY;	PRT;	3011 AA.
ID	O36609	AC		
DT	01-JAN-1998	(T-EMBLrel. 05, Created)		
DT	01-JAN-1998	(T-EMBLrel. 05, Last sequence update)		
DT	01-MAR-2004	(T-EMBLrel. 26, Last annotation update)		
DE	Polyprotein.			
OS	Hepatitis C virus strain H77.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.			
OX	NCBI_TaxID=63746;			
RN	[1]_TaxID=63746;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H77;			
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;			
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;			
RT	Transcripts from a single full-length cDNA clone of hepatitis C virus			
RT	are infectious when directly transfected into the liver of a			
RT	chimpanzee.;			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).			
DR	EMBL; AF011752; AAB67037.1; -.			
DR	PIR; A44150; A44150.			
DR	PIR; PQ0804; PQ0804.			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P26664; 1HEI.			
DR	GO; GO:0019028; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006350; P:transcription; IEA.			
DR	GO; GO:0019079; P:viral genome replication; IEA.			
DR	GO; GO:0019087; P:viral transformation; IEA.			
DR	InterPro; IPR00345; CytC_heme_BS.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002522; HCV_capsid.			
DR	InterPro; IPR002521; HCV_core.			
DR	InterPro; IPR002519; HCV_env.			
DR	InterPro; IPR002531; HCV_NS1.			
DR	InterPro; IPR000745; HCV_NS4a.			
DR	InterPro; IPR001490; HCV_NS4b.			
DR	InterPro; IPR002868; HCV_NS5a.			
DR	InterPro; IPR002166; HCV_RdRP.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR009003; Pept_Ser_Cys.			
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.			
DR	InterPro; IPR007095; RNA_pol_DS_PS.			
DR	InterPro; IPR007094; RNA_pol_PSVir.			
DR	Pfam; PF01543; HCV_capsid; 1.			
DR	Pfam; PF01542; HCV_core; 1.			
DR	Pfam; PF01539; HCV_env; 1.			
DR	Pfam; PF01560; HCV_NS1; 1.			
DR	Pfam; PF01538; HCV_NS2; 1.			
DR	Pfam; PF02907; HCV_NS3; 1.			
DR	Pfam; PF01006; HCV_NS4a; 1.			
DR	Pfam; PF01001; HCV_NS4b; 1.			
DR	Pfam; PF01506; HCV_NS5a; 1.			
DR	Pfam; PF00271; Helicase_C; 1.			
DR	Pfam; PF00998; Viral_RdRP; 1.			
DR	SMART; SM00487; DEXDC; 1.			
DR	PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.			
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;				
KW Polyprotein; Transmembrane.				
SQ SEQUENCE 3011 AA; 327262 MW; 10DIC9702CA9B5DC CRC64;				
Query Match 97.9%; Score 3542; DB 2; Length 3011;				
Best Local Similarity 97.2%; Pred. No. 1.4e-239;				
Matches 667; Conservative 9; Mismatches 10; Indels 0; Gaps 0;				
QY	1	M A P I T A Y A Q O T R G L G C I I T S I T G R D K N Q V E G E V I Q I V S T A A O T F L A T C I N G V C W T Y Y H C A	60	
DB	1026	L A P I T A Y A Q O T R G L G C I I T S I T S R D K N Q V E G E V I Q I V S T A T Q T F L A T C I N G V C W T Y Y H C A	1085	
QY	61	G T R T T A S P K G P V I Q M Y T N V D Q L V G W P A P Q G S R S I T P C T C G S S D L Y L V T R H A D V I P V R R R	120	
DB	1086	G T R T T A S P K G P V I Q M Y T N V D Q L V G W P A P Q G S R S I T P C T C G S S D L Y L V T R H A D V I P V R R R	1145	
QY	121	G D S R G S L L S P R I S Y L K G S S G P L L C P A G H A V G I F R A A V C T R G V A K A V D F I P V E N L E T T M	180	
DB	1146	G D S R G S L L S P R I S Y L K G S S G P L L C P A G H A V G L F R A A V C T R G V A K A V D F I P V E N L G T T M	1205	
QY	181	R S P V F T D N S P P V P Q S F O V A H L H A P T G S G K S T K V P A A Y A A Q G Y K V L V L N P S V A A T L G F G	240	
DB	1206	R S P V F T D N P S P P A V P Q S F O V A H L H A P T G S G K S T K V P A A Y A A Q G Y K V L V L N P S V A A T L G F G	1265	
QY	241	A Y M S K A H G D P N I R T G V R T I T T G S P I T Y S T Y K F L A D G C S G G A Y D I I I C D E C H S T D A T S	300	
DB	1266	A Y M S K A H G V D P N I R T G V R T I T T G S P I T Y S T Y K F L A D G C S G G A Y D I I I C D E C H S T D A T S	1325	
QY	301	I L I G I T V L D Q A E T A G A R L V L A T A T P P G S V T V P H P N I E E V A L S T T G E I P F Y G K A I P L E V I	360	
DB	1326	I L I G I T V L D Q A E T A G A R L V L A T A T P P G S V T V S H P N I E E V A L S T T G E I P F Y G K A I P L E V I	1385	
QY	361	K G R H L I F C H S K K K C D E L A A K L V A L G I N A V A Y R G L D V S V I P P I G D V V V V A T D A L M T G Y T	420	
DB	1386	K G R H L I F C H S K K K C D E L A A K L V A L G I N A V A Y R G L D V S V I P T S G D V V V V S T D A L M T G T	1445	
QY	421	G D F S V I D C N T C V T Q V D F S L O P T F T I E T I L P Q D A V S R T O R R G R T G R G K P G I Y R F V A P G	480	
DB	1446	G D F S V I D C N T C V T Q V D F S L O P T F T I E T I L P Q D A V S R T O R R G R T G R G K P G I Y R F V A P G	1505	
QY	481	E R P S G M F D S V L C E C Y D A G C A W Y E L T P A E T T V R L R A Y M N T P G L P V C Q D H L E F W E G V F T G L	540	
DB	1506	E R P S G M F D S V L C E C Y D A G C A W Y E L T P A E T T V R L R A Y M N T P G L P V C Q D H L E F W E G V F T G L	1565	
QY	541	T H I D A H F L S Q T K O S G E N L P Y L V A Y Q T V C A R A Q A P P S W D M W K C L I R L K P T L H G P T P L L	600	
DB	1566	T H I D A H F L S Q T K O S G E N F P Y L V A Y Q T V C A R A Q A P P S W D M W K C L I R L K P T L H G P T P L L	1625	
QY	601	Y R L G A V Q N E I T L T H P V T K Y I M T C M S A D L E V T S T W L V G G V L A A L A A Y C L S T G C W I V G R	660	
DB	1626	Y R L G A V Q N E V T L T H P I T K Y I M T C M S A D L E V T S T W L V G G V L A A L A A Y C L S T G C W I V G R	1685	
QY	661	V L S G K P A I I P D R E V L Y R E F D E M E E C	686	
DB	1686	I V L S G K P A I I P D R E V L Y Q E F D E M E E C	1711	
RESULT 11				
O36610		PRELIMINARY;	PRT;	3011 AA.
ID	O36610	AC		
DT	01-JAN-1998	(T-EMBLrel. 05, Created)		
DT	01-JAN-1998	(T-EMBLrel. 05, Last sequence update)		
DT	01-MAR-2004	(T-EMBLrel. 26, Last annotation update)		
DE	Polyprotein.			
OS	Hepatitis C virus strain H77.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.			
OX	NCBI_TaxID=63746;			
RN	[1]_TaxID=63746;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H77;			
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;			

Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
 "Transcripts from a single full-length cDNA clone of hepatitis C virus  
 are infectious when directly transfected into the liver of a  
 chimpanzee";  
 Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).  
 RA EMBL; AF011753; AAB67038.1; -;  
 RT PIR; A4150; A4150.  
 RT PIR; PQ0804; PQ0804.  
 DR PIR; PS0326; PS0326.  
 DR PIR; PS0327; PS0327.  
 DR PIR; PS0328; PS0328.  
 DR HSP; P27958; lHEI.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006350; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR000345; CytC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH N.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_S29.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR InterPro; IPR002518; Pept\_U39\_HCV NS2.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; Transmembrane.  
 SQ SEQUENCE 3011 AA; 327222 MW; 293P91430A0D4067 CRC64;  
 Query Match 97.9%; Score 3542; DB 2; Length 3011;  
 Best Local Similarity 97.4%; Pred. No. 1.4e-239;  
 Matches 667; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 APITAYAOQTGLGCIITSLTGDKQVGEVQIVSTAAOTFLATCINGVCWTVYHGAG 61  
 DB 1027 APITAYAOQTGLGCIITSLTGDKQVGEVQIVSTAAOTFLATCINGVCWTVYHGAG 1086  
 QY 62 TRTIASPKGPVIQMTNVNVDQLVGNPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121  
 DB 1087 TRTIASPKGPVIQMTNVNVDQLVGNPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 1146  
 QY 122 DSRGSLSPRISYIKSGSGGPLICPAGHAGVIFRAAVCTRGKAVKADFIPIVENLTTMR 181  
 DB 1147 DSRGSLSPRISYIKSGSGGPLICPAGHAGVIFRAAVCTRGKAVKADFIPIVENLTTMR 1206

182 SPVFTDNSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 241  
 DB 1207 SPVFTDNSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 1266  
 QY 242 YMSKAGHIDPNIRTVRTITGSPITVSTYVKGFLADGCGSGGAYDIIICDECHSTDATSI 301  
 DB 1267 YMSKAGHIDPNIRTVRTITGSPITVSTYVKGFLADGCGSGGAYDIIICDECHSTDATSI 1326  
 QY 302 LGIGTVLDQAETAGARLVVLATATPPGSSVTVPHNIEEVALSTTGEIPFYGKAIPLEVIK 361  
 DB 1327 LGIGTVLDQAETAGARLVVLATATPPGSSVTVPHNIEEVALSTTGEIPFYGKAIPLEVIK 1386  
 QY 362 GGRHLIFCHSKKCKDELAALKVALGINAVAYRGLDVSVPPIGDVWVVAATDALMTGYTG 421  
 DB 1387 GGRHLIFCHSKKCKDELAALKVALGINAVAYRGLDVSVPPIGDVWVVAATDALMTGYTG 1446  
 QY 422 DFDSDVDCNCTVQTQVDFSLDPTFTTITLTPQDASVRSRTORRGTRGKPGIYRFVAPGE 481  
 DB 1447 DFDSDVDCNCTVQTQVDFSLDPTFTTITLTPQDASVRSRTORRGTRGKPGIYRFVAPGE 1506  
 QY 482 RPSGMFDSVLCCEYDAGCAWYELTPTAETTVRLRAYMNTPLGVPQDHLFEWEGVFTGLT 541  
 DB 1507 RPSGMFDSVLCCEYDAGCAWYELTPTAETTVRLRAYMNTPLGVPQDHLFEWEGVFTGLT 1566  
 QY 542 HIDAHLFSQTKSGENLPYLVAQVATVCARAQAPPPSWDQWKKLIRLKLHGTPTLLY 601  
 DB 1567 HIDAHLFSQTKSGENLPYLVAQVATVCARAQAPPPSWDQWKKLIRLKLHGTPTLLY 1626  
 QY 602 RLGAQVONEITLTHPVTKYIMTCHSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGRV 661  
 DB 1627 RLGAQVONEITLTHPVTKYIMTCHSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGRV 1686  
 QY 662 VLSGKPAIIPDREVLVREFDEMEEC 686  
 DB 1687 VLSGKPAIIPDREVLVREFDEMEEC 1711  
 RESULT 12  
 Q03463 ID Q03463 PRELIMINARY; PRT; 3011 AA.  
 AC Q03463; 1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92044440; PubMed=1658196;  
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,  
 RA Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions";  
 RL J. Gen. Virol. 72:2697-2704 (1991).  
 RL (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93117120; PubMed=1335573;  
 RA Okamoto H., Kanai N., Mishihiro S.;  
 RT "Full-length nucleotide sequence of a Japanese hepatitis C virus  
 RT isolate (HC-J1) with high homology to USA isolates";  
 RL Nucleic Acids Res. 20:6410-6410 (1992).  
 RL (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91013116; PubMed=2170712;  
 RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,  
 RA Yoshizawa H., Tada F., Miyakawa Y., Mayumi M.;  
 RT "The 5'-terminal sequence of the hepatitis C virus genome";  
 RL Jpn. J. Exp. Med. 60:167-177 (1990).

[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94174722; PubMed=7510436;  
RA Mink M., Benichou S., Madale P., Tiollais P., Prince A.,  
Inchauspe G.;  
RT "Characterization and mapping of a B-cell immunogenic domain in  
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";  
RL Virology 200:246-255(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Okamoto H.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D10749; BAA01582.1; -;  
DR PIR: PS0326; PS0326.  
DR PIR: PS0327; PS0327.  
DR PIR: PS0328; PS0328.  
DR PIR: S40770; S40770.  
DR HSSP: P26664; 1HEI.  
DR GO: GO:0016021; C:integral to membrane; IEA.  
DR GO: GO:0019028; C:viral capsid; IEA.  
DR GO: GO:0019031; C:viral envelope; IEA.  
DR GO: GO:0005524; F:ATP binding; IEA.  
DR GO: GO:008026; F:ATP-dependent helicase activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO: GO:008236; F:serine-type peptidase activity; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR GO: GO:006508; F:proteolysis and peptidolysis; IEA.  
DR GO: GO:006350; P:transcription; IEA.  
DR GO: GO:0019079; P:viral genome replication; IEA.  
DR GO: GO:0019087; P:viral transformation; IEA.  
DR InterPro: IPR000345; CytC\_heme\_BS.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR011545; DEAD/DEAH\_N.  
DR InterPro: IPR002522; HCV capsid.  
DR InterPro: IPR002521; HCV core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV NS1.  
DR InterPro: IPR000745; HCV NS4a.  
DR InterPro: IPR001490; HCV NS4b.  
DR InterPro: IPR002868; HCV NS5a.  
DR InterPro: IPR002166; HCV NS5b.  
DR InterPro: IPR001650; Helicase C.  
DR InterPro: IPR004109; Peptidase S29.  
DR InterPro: IPR009003; Pept\_Ser\_Cys.  
DR InterPro: IPR002518; Pept\_U3\_HCV\_NS2.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV capsid; 1.  
DR Pfam: PF01542; HCV core; 1.  
DR Pfam: PF01539; HCV env; 1.  
DR Pfam: PF01560; HCV NS1; 1.  
DR Pfam: PF01538; HCV NS2; 1.  
DR Pfam: PF02907; HCV NS3; 1.  
DR Pfam: PF01006; HCV NS4a; 1.  
DR Pfam: PF01001; HCV NS4b; 1.  
DR Pfam: PF01506; HCV NS5a; 1.  
DR Pfam: PF00271; Helicase C; 1.  
DR Pfam: PF00998; Viral RdRp; 1.  
DR SMART: SM00487; DEXDC; 1.  
DR PROSITE: PS00190; CVTOCHROME\_C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327114 MW; 97E9052C0250463B CRC64;  
  
Query Match 97.9%; Score 3542; DB 2; Length 3011;  
Best Local Similarity 96.9%; Pred. No. 1.4e-239;  
Matches 665; Conservative 10; Mismatches 11; Indels 0; Gaps 0;  
  
Qy 1 MAPITAAQTRGLGCIITSLTRDRKNQVGEVQIVSTAAQTFLATCINGCVCTVYHGA 60  
Db :|||||  
1026 LAPITAAQTRGLGCIITSLTRDRKNQVGEVQIVSTAAQTFLATCINGCVCTVYHGA 1085

QY 61 GTRTTIASKPGPIQMYTNVDQDLVGPAPQGGSRSLTPTCTGSSDLYLVTRHADVIPVRR 120  
Db |||||  
1086 GTRTTIASKPGPIQMYTNVDQDLVGPAPQGGARSLTPTCTGSSDLYLVTRHADVIPVRR 1145  
  
QY 121 GDSRGLSPRPISYLKGSSGGLPCPAGHAGIIFRAAVCTRGVAKAVDFIPVENLETMM 180  
Db |||||  
1146 GDSRGLSPRPISYLKGSSGGLPCPAGHVGIIIFRAAVCTRGVAKAVDFIPVESLETMM 1205  
  
QY 181 RSPVFTDNSSPPVPQSFQVAHLHAPTGSCKTKVPAAYAAQGYKVLVLPNSVAATLGF 240  
Db |||||  
1206 RSPVFTDNSSPPAVPQSFQVAHLHAPTGSCKTKVPAAYAAQGYKVLVLPNSVAATLGF 1265  
  
QY 241 AYMSKAHGIDPNIRGTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDTDS 300  
Db |||||  
1266 AYMSKAHGIDPNIRGTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDTDS 1325  
  
QY 301 ILGIGTVLDOAETAGARLVLTATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360  
Db :|||||  
1326 VLGIGTVLDOAETAGARLVLTATATPPGSIIVPHANIEEVALSTTGEIPFYGKAIPLEAI 1385  
  
QY 361 KGRHLIFCHSKKCKDELAALVALGINAVYRGIDVSVIPPIGDVWVWVATDALMTGYT 420  
Db |||||  
1386 KGRHLIFCHSKKCKDELAALVALGINAVYRGIDVSVIPTSGDVWVWVATDALMTGYT 1445  
  
QY 421 GPDFSVIDNCVTQTVDVSLDPTFTIETITLPQDAVSTQRGRTRGRKPGIYRFVAPG 480  
Db |||||  
1446 GPDFSVIDNCVTQTVDVSLDPTFTIETITLPQDAVSTQRGRTRGRKPGIYRFVAPG 1505  
  
QY 481 ERPSGMFDSVLCCEVDACAWYELTPAETTVRLRAYMNTPGIPVCDHLEFWEVFTGL 540  
Db |||||  
1506 ERPSGMFDSVLCCEVDTCAWYELTPAETTVRLRAYMNTPGIPVCDHLEFWEVFTGL 1565  
  
QY 541 THIDAHFLSQTQSGENLPVYAYQATVCARAAQPPSDQMWKCLIRLKPTLHGPTPLL 600  
Db |||||  
1566 THIDAHFLSQTQSGENFPYVAYQATVCARAAQPPSDQMWKCLIRLKPTLHGPTPLL 1625  
  
QY 601 YRLGAVQNEITLTHPVTKYIMTSCMSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGR 660  
Db |||||  
1626 YRLGAVQNEITLTHPVTKYIMTSCMSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGR 1685  
  
QY 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db :|||||  
1686 IVLSGKPAIIPDREVLYREFDEMEEC 1711  
  
RESULT 13  
Q9DIT6  
ID Q9DIT6 PRELIMINARY; PRT; 3011 AA.  
AC Q9DIT6  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Genomic RNA for polyprotein gene.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21014672; PubMed=11115058;  
RA Kumar U., Tuthill T., Thomas H.C., Monjardino J.;  
RT "Sequence, expression and reconstruction of an HCV genome from a  
RT British isolate derived from a single blood donation.";  
RL J. Viral Hepat. 7:459-465(2000).  
DR EMBL: AJ278830; CAC03609.1; -;  
DR PIR: PS0326; PS0326.  
DR PIR: PS0327; PS0327.  
DR PIR: PS0328; PS0328.  
DR HSSP: P27958; 1A1V.  
DR GO: GO:0016021; C:integral to membrane; IEA.  
DR GO: GO:0019028; C:viral capsid; IEA.  
DR GO: GO:0019031; C:viral envelope; IEA.  
DR GO: GO:0005524; F:ATP binding; IEA.

DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
 DR GO: 0003723; F: RNA binding; IEA.  
 DR GO: 0003688; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 0008236; F: serine-type peptidase activity; IEA.  
 DR GO: 0005198; F: structural molecule activity; IEA.  
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: 0006350; P: transcription; IEA.  
 DR GO: 0019079; P: viral genome replication; IEA.  
 DR GO: 0019087; P: viral transformation; IEA.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD\_DEAD\_N.  
 DR InterPro: IPR011545; DEAD\_DEAD\_N.  
 DR InterPro: IPR011492; Flavi\_DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_S29.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR InterPro: IPR002518; Pept\_U39\_HCV\_NS2.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF07652; Flavi\_DEAD; 1.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS0190; CYTOCHROME\_C; UNKNOWN 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Polyprotein; Transmembrane.  
 FT CHAIN 810 1026 non-structural protein 2.  
 FT CHAIN 1027 1657 non-structural protein 3.  
 FT CHAIN 1658 1711 non-structural protein 4a.  
 FT CHAIN 2 191 core protein.  
 FT CHAIN 1712 1972 non-structural protein 4b.  
 FT CHAIN 1973 2420 non-structural protein 5a.  
 FT CHAIN 2421 3011 non-structural protein 5b.  
 FT CHAIN 192 383 envelop protein 1.  
 FT CHAIN 384 809 envelop protein 2.  
 SQ SEQUENCE 3011 AA; 327406 MW; 786264A7A5452D3 CRC64;  
 Query Match 97.8%; Score 3540; DB 2; Length 3011;  
 Best Local Similarity 97.1%; Pred. No. 1.9e-239;  
 Matches 666; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MAPITAAQTRGLGCIITSLTGDRKNQVEQIVSTAQTFLATCINGCVTVYHGA 60  
 DB LPTITAAQTRGLGCIITSLTGDRKNQVEQIVSTAQTFLATCINGCVTVYHGA 1085  
 QY 61 GTRTASPKGPVQNTWVDQDLVGPAPQGRSLTPTCTGSSDLYLVTRHADVIPVRRR 120  
 DB GTRTASPKGPVQNTWVDQDLVGPAPQGRSLTPTCTGSSDLYLVTRHADVIPVRRR 1145  
 QY 121 GDSRGLSPRPISYLKSGSGGLPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
 DB GDGRGLSPRPISYLKSGSGGLPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
 QY 181 RSPVFTDSSPPVPQSQVQVLAHLHAPTSGSKTKVPAAYAGQKVLVNLNSVAATLGF 240  
 DB RSPVFTDSSPPVPQSQVQVLAHLHAPTSGSKTKVPAAYAGQKVLVNLNSVAATLGF 1265

QY 241 AYMSKAHGIDPNIRTVRTITTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSDATS 300  
 DB AYMSKAHGIDPNIRTVRTITTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSDATS 1325  
 QY 301 ILGIGTVLDQAEATAGARLVVLTATATPPGSTVTPHPNIEEVALSTTGIPFVGKAIPLLEVI 360  
 DB ILGIGTVLDQAEATAGARLVVLTATATPPGSTVTPHPNIEEVALSTTGIPFVGKAIPLLEVI 1385  
 QY 361 KGRHLIFCHSKKKCDLAAKVALGINAVAYVYGLDVSVIPPIGDVVVATDALMTGYT 420  
 DB KGRHLIFCHSKKKCDLAAKVALGINAVAYVYGLDVSVIPPIGDVVVATDALMTGYT 1445  
 QY 421 GDFDSVIDCNTCTVTVDFSLDPTFTIETTLTQDAVSRTQRRGRTGRGKGIYRFVAPG 480  
 DB GDFDSVIDCNTCTVTVDFSLDPTFTIETTLTQDAVSRTQRRGRTGRGKGIYRFVAPG 1505  
 QY 481 ERPSGMFDSVLCVCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFWEVGTGL 540  
 DB ERPSGMFDSVLCVCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFWEVGTGL 1565  
 QY 541 THIDAFSLQTSKGENLPYLVAQVATVCARAQPPSWDMKCLIRLKLPTLHGTPPLL 600  
 DB THIDAFSLQTSKGENLPYLVAQVATVCARAQPPSWDMKCLIRLKLPTLHGTPPLL 1625  
 QY 601 YRLGAVONEITLTHPVTKYITMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVWIVGR 660  
 DB YRLGAVONEITLTHPVTKYITMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVWIVGR 1685  
 QY 661 VVLSGKPAIIPDREVLVREFDEMEEC 686  
 DB IVLSGKRPPIIPDREVLVREFDEMEEC 1711  
 RESULT 14  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID 01-AUG-1992 (Rel. 23, Created)  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE NSB (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467 (1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270; DOI=10.1016/S0969-2126(98)00010-0;  
 RA Kim J.L., Morgerstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode



```
Qy 61 GTRTIASPKGPVIOYNTNVDDLVGWPAPOGSRSLTPTCTCGSSDLYLVTTRHADVIPVRRR 120
D 1086 GTRTIASPKGPVIOYNTNVDDLVGWPAPOGSRSLTPTCTCGSSDLYLVTTRHADVIPVRRR 1145
Qy 121 GDSRGSLLSPPIISYLVKSGSGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 180
D 1146 GDSRGSLLSPPIISYLVKSGSGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 1205
Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAQAGYKVLVLPSPVAATLGF 240
D 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAQAGYKVLVLPSPVAATLGF 1265
Qy 241 AYMSKAHGIDNIRTVGRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATS 300
D 1266 AYMSKAHGIDNIRTVGRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATS 1325
Qy 301 ILGIGTVLDQAEAGARLVLATATPGSVTVPHNTEEVALSTGEIPFYGKAIPLEVI 360
D 1326 ISGIGTVLDQAEAGARLVLATATPGSVTVPHNTEEVALSTGEIPFYGKAIPLEVI 1385
Qy 361 KGGRHLLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALMTGYT 420
D 1386 KGGRHLLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALMTGYT 1445
Qy 421 GDFDSVIDCNTCVTQTVDFSLDPTFTTITLTPQDAVSRTQRRGTRGKPGIYRFVAPG 480
D 1446 GDFDSVIDCNTCVTQTVDFSLDPTFTTITLTPQDAVSRTQRRGTRGKPGIYRFVAPG 1505
Qy 481 ERPSGMFDDSVLCBICYDAGCAWYELTFAETTVRLRAYMNTPLGVCQDHLGFWGVFTGL 540
D 1506 ERPSGMFDDSVLCBICYDAGCAWYELTFAETTVRLRAYMNTPLGVCQDHLGFWGVFTGL 1565
Qy 541 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSWDMWKLRLKPLTLAGPTPLL 600
D 1566 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSWDMWKLRLKPLTLAGPTPLL 1625
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWVI 660
D 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWVI 1685
Qy 661 VVLGKPAIIPDREVLYREFDEMEEC 686
D 1686 IVLGGKPAIIPDREVLYREFDEMEEC 1711
RESULT 15
Q9DTE2 PRELIMINARY; PRT; 3010 AA.
AC Q9DTE2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Serum;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RL Hepatol. Res. 20:161-171(2001).
DR EMBL; AB049095; BAB18808.1; -.
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00252; P00252.
DR PIR; P00253; P00253.
DR PIR; P00254; P00254.
```

```
DR PIR; P00255; P00255.
DR PIR; P00804; P00804.
DR PIR; P0329; P0329.
DR HGSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005244; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:aerine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD_DEAD_N.
DR InterPro; IPR011545; DEAD_DEAD_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid_1.
DR Pfam; PF01542; HCV_core_1.
DR Pfam; PF01539; HCV_core_1.
DR Pfam; PF01560; HCV_NS1_1.
DR Pfam; PF01538; HCV_NS2_1.
DR Pfam; PF02907; HCV_NS3_1.
DR Pfam; PF01006; HCV_NS4a_1.
DR Pfam; PF01001; HCV_NS4b_1.
DR Pfam; PF01506; HCV_NS5a_1.
DR Pfam; PF00271; Helicase_C_1.
DR Pfam; PF00998; Viral_RdRP_1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3010 AA; 326868 MW; 25BE54B9D7EEA15 CRC64;
Query Match 94.9%; Score 3436; DB 2; Length 3010;
Best Local Similarity 92.6%; Pred. No. 3.9e-232;
Matches 635; Conservative 34; Mismatches 17; Indels 0; Gaps 0;
Qy 1 MAPITAYAQTRGLGCIITSLTRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 60
D 1026 LAPITAVSQTRGLGCIITSLTRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 1085
Qy 61 GTRTIASPKGPVIOYNTNVDDLVGWPAPOGSRSLTPTCTCGSSDLYLVTTRHADVIPVRRR 120
D 1086 GSKTLAGPKGPITQMTNVDDLVGWPAPOGSRSLTPTCTCGSSDLYLVTTRHADVIPVRRR 1145
Qy 121 GDSRGSLLSPPIISYLVKSGSGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 180
D 1146 GDSRGSLLSPPIISYLVKSGSGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMM 1205
Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAQAGYKVLVLPSPVAATLGF 240
D 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAQAGYKVLVLPSPVAATLGF 1265
Qy 241 AYMSKAHGIDNIRTVGRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATS 300
D 1266 AYMSKAHGIDNIRTVGRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATS 1325
```



Db	1266	AYMSKAHVDPNIRTVRTITTCAPITYSTYKFLADGGCGGAYDIIICDECHSTDSTS	1325
Qy	301	ILGIGTVLQQAETAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIIPFYGKAIPELEVI	360
Db	1326	ILGIGTVLQQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIIPFYGKAIPIETI	1385
Qy	361	KGGRHLIFCHSKKKCDLAAKLVALGINAVAYYRGLDVSIPPIGDVVVVVATDALMTGYT	420
Db	1386	KGGRHLIFCHSKKKCDLAAKLVALGVNAVAYYRGLDVSIPPTSGDVVVVATDALMTGYT	1445
Qy	421	GPDFSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTQREGRTGRGKPGIYRFVAPG	480
Db	1446	GPDFSVIDCNTCVTQTVDFSLDPTFTIETITVPQDAVSRSQRRGRTGRGGIYRFVTPG	1505
Qy	481	ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFWEGVFTGL	540
Db	1506	ERPSGMFDSVLCCEYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLEFWESVFTGL	1565
Qy	541	THIDAHFLSOTKOSGENLPYLVAIQATVCARAQAPPPSWDMWKCLIRLKPTLHGPTPLL	600
Db	1566	THIDAHFLSOTKQAGDNFPYLVAIQATVCARAQAPPPSWDMWKCLIRLKPTLHGPTPLL	1625
Qy	601	YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTGCWVIVGR	660
Db	1626	YRLGAVQNDVTLTHPVTKFIMACMSADLEVVTSTWLVGGVLAALAAAYCLTSGSVIVGR	1685
Qy	661	VVLSGKPAIIPDREVLYREFDEMEEC	686
Db	1686	IILSGRPALIPDREVLYQEFDEMEEC	1711

Search completed: November 7, 2005, 20:16:35  
Job time : 91.9151 secs

Result No.	Query Match	Score	Length	DB	ID	Description
1	4455	100.0	829	5	AAE18690	AAE18690 Multiple
2	4455	100.0	829	7	ADC06769	ADC06769 Chimeric
3	4455	100.0	829	8	ADL66807	ADL66807 HCV multi
4	4032	90.5	1099	5	AAU76378	AAU76378 HCV multi
5	4032	90.5	1099	6	ABG72262	ABG72262 HCV multi
6	4032	90.5	1099	8	ADL66809	ADL66809 HCV multi
7	2222	49.9	1021	2	AAW34481	AAW34481 HCV anti
8	2222	49.9	1021	2	AAW40039	AAW40039 Fusion pr
9	2222	49.9	1021	5	AAE22050	AAE22050 pSOD/c200
10	1624.5	36.5	2261	1	AAE32041	AAE32041 Hepatitis
11	1624.5	36.5	2261	1	AAE32041	AAE32041 Hepatitis
12	1624.5	36.5	2436	1	AAE32050	AAE32050 HCV prote
13	1624.5	36.5	2436	1	AAE32050	AAE32050 HCV prote
14	1624.5	36.5	2772	3	AAAB18540	AAAB18540 Protein e
15	1624.5	36.5	2772	8	ADN35976	ADN35976 HCV cdNA
16	1624.5	36.5	2955	2	AAV14975	AAV14975 Amino aci
17	1624.5	36.5	2955	3	AAAB18541	AAAB18541 Polyprote
18	1624.5	36.5	2955	8	ADN35978	ADN35978 HCV cdNA
19	1624.5	36.5	3011	2	AAAR21519	AAAR21519 Compel
20	1624.5	36.5	3011	2	AAAR30931	AAAR30931 Hepatitis
21	1624.5	36.5	3011	2	AAW34480	AAW34480 HCV poly
22	1624.5	36.5	3011	2	AAW40038	AAW40038 HCV poly
23	1624.5	36.5	3011	5	AAE22049	AAE22049 Hepatitis
24	1624.5	36.5	3011	8	ADL23107	ADL23107 Hepatitis
25	1624.5	36.5	3011	8	ADR29357	ADR29357 Hepatitis

CC matrix as well as immunoassay solid supports for use in the assay. The  
CC solid support is useful for detecting HCV infection in a biological  
CC sample. The present sequence is MEFA (multiple epitope fusion antigen) 12  
CC protein. This sequence is used in the exemplification of the invention  
XX  
SQ Sequence 829 AA;

Query Match 100.0%; Score 4455; DB 5; Length 829;  
Best Local Similarity 100.0%; Pred. No. 2.4e-310;  
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Db 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Qy 121 AARTTSGLTSLFSPGASONIQLITSTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Db 121 AARTTSGLTSLFSPGASONIQLITSTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Qy 181 AQGYKVLVLPNPSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYTYGKFLADGGC 240  
Db 181 AQGYKVLVLPNPSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYTYGKFLADGGC 240  
Qy 241 SGGAYDIIICDECHSTATSIIGTGLTVDQABTAGARLVLTATPPGTVTPHPNIEEV 300  
Db 241 SGGAYDIIICDECHSTATSIIGTGLTVDQABTAGARLVLTATPPGTVTPHPNIEEV 300  
Qy 301 ALSTTGIEIPFYGKAIPLVETKGGHLLFCHSKKCDLAAKLVALGINAVAYYRGLDVSV 360  
Db 301 ALSTTGIEIPFYGKAIPLVETKGGHLLFCHSKKCDLAAKLVALGINAVAYYRGLDVSV 360  
Qy 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTACSGKPAIIPDREVLRYREDEMECSQH 420  
Db 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTACSGKPAIIPDREVLRYREDEMECSQH 420  
Qy 421 LPYIEQGMMLAEQPKQKALGLSRGKPAIVDPKEVLVQOYDEMECSQAAPYIEQAQVIA 480  
Db 421 LPYIEQGMMLAEQPKQKALGLSRGKPAIVDPKEVLVQOYDEMECSQAAPYIEQAQVIA 480  
Qy 481 HQPEKVLGLIDNDQVVVTPDKETLYEAFDMECSKAAALIEGQRMALMSKIOGLL 540  
Db 481 HQPEKVLGLIDNDQVVVTPDKETLYEAFDMECSKAAALIEGQRMALMSKIOGLL 540  
Qy 541 GILRRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPNPLV 600  
Db 541 GILRRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPNPLV 600  
Qy 601 ETWKKPDYEPVHVGRSSRRFAQALPVWARPDPNPLVETWKKPDYEPVHVGRKTKRNT 660  
Db 601 ETWKKPDYEPVHVGRSSRRFAQALPVWARPDPNPLVETWKKPDYEPVHVGRKTKRNT 660  
Qy 661 NRRPDQVFKPGGQIVGVYLLPRRGLVLATRTKTSPIPKARRPREGRTWAQPGYWPPL 720  
Db 661 NRRPDQVFKPGGQIVGVYLLPRRGLVLATRTKTSPIPKARRPREGRTWAQPGYWPPL 720  
Qy 721 YGNKDRSTGKSWGKPGYWPMPKRTKRNTRRQDVQKFPGGQIVGVYLLPRRGLVL 780  
Db 721 YGNKDRSTGKSWGKPGYWPMPKRTKRNTRRQDVQKFPGGQIVGVYLLPRRGLVL 780  
Qy 781 ATRKTSIPKARRPREGRTWAQPGYWPPLYGNKDRSTGKSWGKPGYWP 829  
Db 781 ATRKTSIPKARRPREGRTWAQPGYWPPLYGNKDRSTGKSWGKPGYWP 829

RESULT 2  
ADC06769  
ID ADC06769 standard; protein; 829 AA.  
XX  
AC ADC06769;

XX 18-DEC-2003 (first entry)  
DT Chimeric multiple epitope fusion antigen 12 protein.  
DE  
XX immunoassay solid support; HCV; NS3/4a; non-structural;  
KW non-A, non-B hepatitis; NANB; multiple epitope fusion antigen 12; MEFA12;  
KW chimeric.  
XX Chimeric.  
OS Synthetic.  
OS Unidentified.  
OS Hepatitis C virus.  
OS Homo sapiens.  
XX US2002192639-A1.  
XX 19-DEC-2002.  
XX 14-JUN-2001; 2001US-00881239.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIE/) CHIEN D Y.  
PA (ARCA/) ARCANGEL P.  
PA (TAND/) TANDESKE L.  
PA (GEOR/) GEORGE-NASCIMENTO C.  
PA (COIT/) COIT D.  
PA (MEDI/) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX WPI; 2003-644609/61.  
DR N-PSDB; ADC06770.  
XX  
PT Immunoassay solid support for detecting hepatitis C virus infection in  
PT biological samples, comprises a hepatitis C virus anti-core antibody and  
PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
PT antibody.  
XX Claim 45; Fig 7; 40pp; English.  
XX The invention relates to a novel immunoassay solid support comprising at  
CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereco.  
CC The system of the invention may be useful for detecting HCV infection in  
CC a biological sample and for treating or detecting non-A, non-B hepatitis  
CC (NANB hepatitis). The current sequence is that of the chimeric multiple  
CC epitope fusion antigen 12 (MEFA12) protein of the invention.  
XX  
SQ Sequence 829 AA;  
Query Match 100.0%; Score 4455; DB 7; Length 829;  
Best Local Similarity 100.0%; Pred. No. 2.4e-310;  
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Db 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Qy 121 AARTTSGLTSLFSPGASONIQLITSTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Db 121 AARTTSGLTSLFSPGASONIQLITSTDNSSPPVPPQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Qy 181 AQGYKVLVLPNPSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYTYGKFLADGGC 240  
XX

Db 181 AQQYKVLNPSVAATLGFQYMSKAGHIDPNIRGTGRTITTTGSPITTYTGKFLADGGC 240  
 Qy 241 SGGAYDIIICDECHSTDATSLTGITGVLDOAETAGARLVVLATATPPGVTVPHPNIEEV 300  
 Db 241 SGGAYDIIICDECHSTDATSLTGITGVLDOAETAGARLVVLATATPPGVTVPHPNIEEV 300  
 Qy 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS 360  
 Db 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS 360  
 Qy 361 IPTSGDVVVATDALMTGYTGDVSDVDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
 Db 361 IPTSGDVVVATDALMTGYTGDVSDVDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
 Qy 421 LPYIEQGMMLAEQFKQKALGLSRGKPAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
 Db 421 LPYIEQGMMLAEQFKQKALGLSRGKPAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
 Qy 481 HQFKEKVLGLINDOVVVTDPKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
 Db 481 HQFKEKVLGLINDOVVVTDPKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
 Qy 541 GILRRHVGPGEVQVMNRLIAFASRGNHVSPTHVPSRRFAQALPVWARPDPNPLV 600  
 Db 541 GILRRHVGPGEVQVMNRLIAFASRGNHVSPTHVPSRRFAQALPVWARPDPNPLV 600  
 Qy 601 ETWKKPDYEPVPHVGRSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHVGRKTKRNT 660  
 Db 601 ETWKKPDYEPVPHVGRSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHVGRKTKRNT 660  
 Qy 661 NRRPDVKFPGGGQIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRWQPGYWPPL 720  
 Db 661 NRRPDVKFPGGGQIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRWQPGYWPPL 720  
 Qy 721 YGNKDRSTGKSGKPGYPWPRKTKRNTNRRPDVKFPGGGQIVGGVYLLPRGPRGLV 780  
 Db 721 YGNKDRSTGKSGKPGYPWPRKTKRNTNRRPDVKFPGGGQIVGGVYLLPRGPRGLV 780  
 Qy 781 ATRKTSPIPKARRPEGRWQPGYWPWPLYGNKDRRSTGKSGKPGYPWP 829  
 Db 781 ATRKTSPIPKARRPEGRWQPGYWPWPLYGNKDRRSTGKSGKPGYPWP 829

RESULT 3  
 ADL66807  
 ID ADL66807 standard; protein; 829 AA.  
 XX  
 AC ADL66807;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE HCV multiple epitope fusion antigen 12 (MEFA 12) polypeptide.  
 XX  
 KW HCV; MEFA 12; HCV antigen; HCV polyprotein;  
 KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection;  
 KW multiple epitope fusion antigen 12.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PN WO2004021871-A2.  
 XX  
 PD 18-MAR-2004.  
 XX  
 PF 08-SEP-2003; 2003WO-US028071.  
 XX  
 PR 09-SEP-2002; 2002US-0409515P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Arcangel P, Chien D;  
 XX  
 DR WPI; 2004-248333/23.  
 DR N-PSDB; ADL66806.

XX  
 PT Detecting hepatitis C virus (HCV) infection in a biological sample by  
 PT detecting complexes formed between the HCV antibody and the antigens from  
 PT the first region of the HCV polyprotein and the multiple epitope fusion  
 PT antigen (MEFA).  
 XX  
 PS Claim 14; SEQ ID NO 4; 93pp; English.  
 XX  
 CC The invention relates to a method of detecting hepatitis C virus (HCV)  
 CC infection in a biological sample. The method comprises providing an  
 CC immunassay solid support comprising HCV antigens bound to it, where the  
 CC HCV antigens comprise one or more isolated antigens form a first region  
 CC of the HCV polyprotein, combining a biological sample with the solid  
 CC support under conditions that allow HCV antibodies, when present in the  
 CC biological sample, to bind to the one or more HCV antigens, adding to the  
 CC solid support a detectably labelled HCV multiple epitope fusion antigen  
 CC (MEFA), where the labelled MEFA comprises at least one epitope from the  
 CC same region of the HCV polyprotein as the one or more isolated antigens,  
 CC where the MEFA binds to the bound HCV antibody, and detecting complexes  
 CC formed between the HCV antibody and the one or more antigens from the  
 CC first region of the HCV polyprotein and the MEFA, if any, as an  
 CC indication of HCV infection in the biological sample. The method is  
 CC useful for detecting hepatitis C virus (HCV) infection in a biological  
 CC sample. This sequence represents the MEFA 12 polypeptide used in the  
 CC scope of the invention.  
 XX  
 SQ Sequence 829 AA;  
 Query Match 100.0%; Score 4455; DB 8; Length 829;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-310;  
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MATKAVCVLKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
 Db 1 MATKAVCVLKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
 Qy 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWLKLSAARTTSGFVSLFAPCAKQNETHVTTGA 120  
 Db 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWLKLSAARTTSGFVSLFAPCAKQNETHVTTGA 120  
 Qy 121 AARTTSGTSLFSPGASQNIQLITSTDNSSPPVPPQSFQVAHLHAPTSGSKTKVPAAYA 180  
 Db 121 AARTTSGTSLFSPGASQNIQLITSTDNSSPPVPPQSFQVAHLHAPTSGSKTKVPAAYA 180  
 Qy 181 AQQYKVLNPSVAATLGFQYMSKAGHIDPNIRGTGRTITTTGSPITTYTGKFLADGGC 240  
 Db 181 AQQYKVLNPSVAATLGFQYMSKAGHIDPNIRGTGRTITTTGSPITTYTGKFLADGGC 240  
 Qy 241 SGGAYDIIICDECHSTDATSLTGITGVLDOAETAGARLVVLATATPPGVTVPHPNIEEV 300  
 Db 241 SGGAYDIIICDECHSTDATSLTGITGVLDOAETAGARLVVLATATPPGVTVPHPNIEEV 300  
 Qy 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS 360  
 Db 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS 360  
 Qy 361 IPTSGDVVVATDALMTGYTGDVSDVDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
 Db 361 IPTSGDVVVATDALMTGYTGDVSDVDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
 Qy 421 LPYIEQGMMLAEQFKQKALGLSRGKPAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
 Db 421 LPYIEQGMMLAEQFKQKALGLSRGKPAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
 Qy 481 HQFKEKVLGLINDOVVVTDPKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
 Db 481 HQFKEKVLGLINDOVVVTDPKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
 Qy 541 GILRRHVGPGEVQVMNRLIAFASRGNHVSPTHVPSRRFAQALPVWARPDPNPLV 600  
 Db 541 GILRRHVGPGEVQVMNRLIAFASRGNHVSPTHVPSRRFAQALPVWARPDPNPLV 600  
 Qy 601 ETWKKPDYEPVPHVGRSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHVGRKTKRNT 660

|||||  
601 ETWKPKDYEPVHVGRSSRFAQALPVWARPDPYNPPLVETWKKDYEPVHVGRKTKRNT 660  
661 NRRPQDVKFGGQIIVGGVLLPRGRPLGLVATRKTSPIPKARRPGRRTWAQGYWPPL 720  
661 NRRPQDVKFGGQIIVGGVLLPRGRPLGLVATRKTSPIPKARRPGRRTWAQGYWPPL 720  
721 YGNKDRRSTGKSGKPGYPWPRKTNRNRRPQDVKFGGQIIVGGVLLPRGRPLGLV 780  
721 YGNKDRRSTGKSGKPGYPWPRKTNRNRRPQDVKFGGQIIVGGVLLPRGRPLGLV 780  
781 ATRKTSPIPKARRPGRRTWAQGYWPPLYGNKDRRSTGKSGKPGYPWP 829  
781 ATRKTSPIPKARRPGRRTWAQGYWPPLYGNKDRRSTGKSGKPGYPWP 829

RESULT 4  
AAU76378  
ID AAU76378 standard; protein; 1099 AA.  
XX  
AC AAU76378;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE HCV multiple epitope fusion antigen (MEFA) 7.1 protein sequence.  
XX  
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
XX immunoassay solid support; multiple epitope fusion antigen; MEFA;  
XX non-structural protein.  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
XX W0200196870-A2.  
XX  
XX 20-DEC-2001.  
XX  
XX 14-JUN-2001; 2001WO-US019156.  
XX  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX  
XX WPI; 2002-090228/12.  
DR N-PSDB; ABK15345.  
XX  
XX Immunoassay solid support, useful for detecting hepatitis C virus  
PT infection in biological sample, comprises HCV NS3/4a conformational  
PT epitope and multiple epitope fusion antigen bound to the support.  
XX  
XX Claim 5; Fig 5; 92pp; English.  
XX  
XX The present invention relates to a new immunoassay solid support  
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a  
CC conformational epitope and a multiple epitope fusion antigen (MEFA),  
CC bound to the support. The NS3/4a conformational epitope and/or MEFA  
CC reacts specifically with anti-HCV antibodies present in a biological  
CC sample from an HCV-infected individual. The immunoassay of the invention  
CC is useful for detecting hepatitis C virus infection in a biological  
CC sample. The method of the invention provides a sensitive, accurate  
CC diagnostic and prognostic tool to provide adequate patient care and to  
CC prevent transmission of HCV by blood and by blood products, or by  
CC personal contact. Use of NS3/4a conformational epitope in combination  
CC with MEFA, provides a sensitive and reliable method for detecting early  
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing  
CC masking problems, improving sensitivity in detecting antibodies by  
CC allowing a greater number of epitopes on a unit surface area of  
CC substrate, and improving substrate. Detection accuracy is increased and

CC the incidence of false results is reduced because of the identification  
CC and the use of highly immunogenic HCV antigens which are present during  
CC the early stages of HCV seroconversion. The present amino acid sequence  
CC represents the multiple epitope fusion antigen (MEFA) 7.1 of the  
CC invention  
XX  
SQ Sequence 1099 AA;  
Query Match 90.5%; Score 4032; DB 5; Length 1099;  
Best Local Similarity 69.8%; Pred. No. 7.6e-280;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESGPKVWVNGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESGPKVWVNGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
Qy 61 AGPHFNPPLSTR----- 71  
Db 61 AGPHFNPPLSRKHGKPKDEHRHVGDLGNVTADKGVADSVIEDSVISLGDHCIIGRTLIV 120  
Qy 72 -----GCNCSIYPGHITGHRMAWKLS 93  
Db 121 HEKADDLGKGNNEESTKTGNAGSRLACGVIGIAQNLNSGCNCSIYPGHITGHRMAWKLS 180  
Qy 94 AARTTSGFVSLFAPGAKQNETHTVGTGAARTTSLTSLFSPGASQNIQLITS----- 145  
Db 181 AARTTSGFVSLFAPGAKQNETHTVGTGAARTTSLTSLFSPGASQNIQLIVDPIPVENLE 240  
Qy 146 -----TNSSPVPVQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPSPAATL 197  
Db 241 TTMRSPPVFTDNSSPPVQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPSPAATL 300  
Qy 198 GFGAYMSKAHGIDPNIRITGVTITTSPTIYSTYTKFLADGGCGGAYDIIICDECHSTD 257  
Db 301 GFGAYMSKAHGIDPNIRITGVTITTSPTIYSTYTKFLADGGCGGAYDIIICDECHSTD 360  
Qy 258 ATSILGIGTVLDDAETAGARLVVLTATATPPGSVTVPHNPTEEVALSTGTBIPFYGKAIP 317  
Db 361 ATSILGIGTVLDDAETAGARLVVLTATATPPGSVTVPHNPTEEVALSTGTBIPFYGKAIP 420  
Qy 318 EVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSTPTSGDVVVVATDALMT 377  
Db 421 EVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSTPTSGDVVVVATDALMT 480  
Qy 378 GYTGDPSVIDCNTC----- 392  
Db 481 GYTGDPSVIDCNTCTQTQVDFSLDPTFTTITLTPQDAVSRQRRGRTGRGKPGIYRFV 540  
Qy 393 ----- 392  
Db 541 APCRPSGMFDSVLCBCEYDAGCAWVELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVF 600  
Qy 393 ----- 392  
Db 601 TGLTHIDAHFLSQTOSGENLPYLVAQATVCARAQAPPSPSQMVKLRLKPLTLHGPT 660  
Qy 393 -----ACSGKPAIIPDREVLRYREFDEME 416  
Db 661 PLIYRLGAVQNEITLTHPTVKYIMTCSADLEVVTSACSKPAIIPDREVLRYREFDEME 720  
Qy 417 CSQHLPIYEGQMLABQFKOKALGLSRGKPKAIPVDPKEVLYQQYDEMECSQAAPYIEQA 476  
Db 721 CSQHLPIYEGQMLAEQFKOKALGLSRGKPKAIPVDPKEVLYQQYDEMECSQAAPYIEQA 780  
Qy 477 QVIAHQFKEKVLGLINDQVVTDPKXILYEAFDEMEECASKALITEGORMAEMLSKI 536  
Db 781 QVIAHQFKEKVLGLINDQVVTDPKXILYEAFDEMEECASKALITEGORMAEMLSKI 840  
Qy 537 QGLLGILRRHVGPGEAVQMMNELLAFASGNHVSPTHYVPSRSRRFAQALPVWARPDPYN 596  
Db 841 QGLLGILRRHVGPGEAVQMMNELLAFASGNHVSPTHYVPSRSRRFAQALPVWARPDPYN 900  
Qy 597 PPLVETWKKPDYBPPFVVHVGSRSSRRFAQALPVWARPDPYNPPLVETWKKPDYBPPFVVHVGKRT 656

Db 901 PPLVETWKEDYBPVHGRSSRRFAQALPVMARPDYNPLVETWKKPDYEPVHGRKT 960  
Qy 657 KRNTNRPDQKPPGGQIVGGVLLPRRGLVLAATKTSPTPKARRPEGRWAOQGY 716  
Db 961 KRNTNRPDQKPPGGQIVG-----RRGP-----PIPKARRPEGRWAOQGY 1003  
Qy 717 PWPLYGNKDRSRSTGSKGKGYPMWRKTKRNTNRRPDQKPPGGQIVGGVLLPRRGR 776  
Db 1004 PWPLYGNKDRSRSTGSKGKGYPMWRKTKRNTNRRPDQKPPGGQIVG-----RRGP- 1056  
Qy 777 LGVLATKTSPIPKARRPEGRWAOQGYPMWRKTKRNTNRRPDQKPPGGQIVGGVLLPRRGR 829  
Db 1057 -----PIPKARRPEGRWAOQGYPMWRKTKRNTNRRPDQKPPGGQIVG-----RRGP- 1099

RESULT 5  
ABG72262  
ID ABG72262 standard; protein; 1099 AA.  
XX ABG72262;  
XX AC ABG72262;  
XX DT 06-MAR-2003 (first entry)  
XX HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).  
XX Immunassay solid support; Hepatitis C Virus type-1; HCV-1; HCV-2;  
XX NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;  
XX MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;  
XX HCV infection; Hepatitis C Virus type-2; Hepatitis C Virus type-3;  
XX mutant; muteln.  
XX Hepatitis C virus type 1.  
XX Hepatitis C virus type 2.  
XX Hepatitis C virus type 3.  
XX Synthetic.  
XX Chimeric.

Key Location/Qualifiers  
Region 1..156  
/note= "Correspond to amino acids 1-156 of HCV-1 HSD  
superoxide dismutase")  
Region 159..176  
/note= "Correspond to amino acids 303-320 of HCV-1 E1"  
Region 179..199  
/note= "Correspond to consensus sequence of amino acids  
390-410 of HCV-1 E2 HVR"  
Region 200..230  
/note= "Correspond to consensus sequence of amino acids  
384-414 of HCV-1 and HCV-2 E2 HVR"  
Region 231..696  
/note= "Correspond to amino acids 1193-1658 of HCV-1  
helicase"  
Region 699..745  
/note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
1 epitope"  
Region 748..794  
/note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
1 epitope"  
Region 797..843  
/note= "Correspond to amino acids 1689-1735 of HCV-2 5-1-  
1 epitope"  
Region 846..881  
/note= "Correspond to amino acids 1901-1936 of HCV-1  
polypeptide C100"  
Region 884..919  
/note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
region"  
Region 922..957  
/note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
region"  
Region 958..1028  
/note= "Correspond to core region antigenic determinants

FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
FT 1029..1099  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
XX  
PN US2002146685-A1.  
XX 10-OCT-2002.  
XX 14-JUN-2001; 2001US-00881654.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIE/) CHIEN D Y.  
PA (ARCA/) ARCANGEL P.  
PA (TAND/) TANDESKE L.  
PA (GEOR/) GEORGE-NASCIMENTO C.  
PA (COIT/) COIT D.  
PA (MEDI/) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
PI WPI; 2003-147573/14.  
XX N-PSDB; ABX14411.  
DR  
XX Immunassay solid support for detecting Hepatitis C Virus infection in  
FT biological samples, comprises Hepatitis C Virus conformational epitope  
FT and multiple epitope fusion antigen.  
XX Claim 25; Fig 5A-5F; 45pp; English.  
XX The present invention relates to immunoassays comprising Hepatitis C  
CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
CC multiple epitope fusion antigen react with anti-HCV antibodies present in  
CC a biological sample from an HCV-infected individual. The immunoassays and  
CC methods of the invention are useful for detecting HCV infection in a  
CC biological sample. The inventive immunoassay solid support provides a  
CC sensitive and reliable method for detecting early HCV seroconversion. The  
CC assays can detect HCV infection caused by any six known genotypes of HCV.  
CC The use of the multiple epitope fusion proteins decreases masking  
CC problems, improves sensitivity in detecting antibodies by allowing a  
CC greater number of epitopes on a unit area of substrate, and improves  
CC selectivity. The present sequence represents HCV multiple epitope fusion  
CC antigen 7.1 (MEFA 7.1), a mutant HCV polypeptide derived from various  
CC regions of HCV type 1, 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide  
XX sequences  
SQ Sequence 1099 AA;  
Query Match 90.5%; Score 4032; DB 6; Length 1099;  
Best Local Similarity 69.8%; Pred. No. 7.6e-280;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTR----- 71  
Db 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCIIIGTLVW 120  
Qy 72 -----GNCSTIYPGHITGHRMAWKLS 93  
Db 121 HEKADDLKKGNEBSTXTGNAGSRLACGVIGIAQLNLSGNCSTIYPGHITGHRMAWKLS 180  
Qy 94 AARTTSFVSLFAPKAKONETHVTGGAAARTTSLGLTSLFSPGASQNTLIITS----- 145

Db	181	AARTTSGFVSLFAPGAKQKQETHVTGGAARTTSGLTSLFSPGASQNIQLIVDFIPVENLE	240	KW	HCV; MEFA 7.1; HCV antigen; HCV polyprotein;
Qy	146	-----TDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL	197	KW	multiple epitope fusion antigen; MEFA; hepatitis C virus infection;
Db	241	TTWRSPTVTDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL	300	XX	multiple epitope fusion antigen 7.1.
Qy	198	GFGAYMSKAHGIDPNIRTTGSPITYSYTGKFLADGGCGSGAYDIICDECHSTD	257	OS	Hepatitis C virus.
Db	301	GFGAYMSKAHGIDPNIRTTGSPITYSYTGKFLADGGCGSGAYDIICDECHSTD	360	PN	WO2004021871-A2.
Qy	258	ATSILGIGTVLQDAETAGARLVVLTATPPGSTVTPHPNIEEVALSTTGEIPYGAIPL	317	XX	18-MAR-2004.
Db	361	ATSILGIGTVLQDAETAGARLVVLTATPPGSTVTPHPNIEEVALSTTGEIPYGAIPL	420	PF	08-SEP-2003; 2003WO-US028071.
Qy	318	EVTKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVIPSGDVVVVATDALMT	377	XX	09-SEP-2002; 2002US-0409515P.
Db	421	EVTKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVIPSGDVVVVATDALMT	480	PR	(CHIR ) CHIRON CORP.
Qy	378	GYTGFDSVIDCNTC-----	392	PA	Arcangel P, Chien D;
Db	481	GYTGFDSVIDCNTCVTQTVDPSLDTFTIETITLPQDAVSRTQRRGTGRGPGIYRFV	540	PI	WPI; 2004-248333/23.
Qy	393	-----	392	DR	N-PSDB; ADL66808.
Db	541	APGERBSGMFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFPWEGVF	600	XX	Detecting hepatitis C virus (HCV) infection in a biological sample by
Qy	393	-----	392	PT	detecting complexes formed between the HCV antibody and the antigens from
Db	601	TGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSQDMWKILRLKPTLHGPT	660	PT	the first region of the HCV polyprotein and the multiple epitope fusion
Qy	393	-----ACSGKPAIIPDREVLRYREFDEME	416	PT	antigen (MEFA).
Db	661	PLLYRLGAVQNETLTHPVTKYIMTCMSADLEVVTACSCKPALIIPDREVLRYREFDEME	720	XX	Claim 15; SEQ ID NO 6; 93pp; English.
Qy	417	CSQHLPIEQQMLABQFKQKALGLSRGKPAIVDPKEVLYQYDEMECSQAAPYIEQA	476	CC	The invention relates to a method of detecting hepatitis C virus (HCV)
Db	721	CSQHLPIEQQMLABQFKQKALGLSRGKPAIVDPKEVLYQYDEMECSQAAPYIEQA	780	CC	infection in a biological sample. The method comprises providing an
Qy	477	QVIAHOFKEKVLGINDQVVVTPDKELIYEADEMEECASKAALIEEGORMAEMLSKI	536	CC	immunoassay solid support comprising HCV antigens bound to it, where the
Db	781	QVIAHOFKEKVLGINDQVVVTPDKELIYEADEMEECASKAALIEEGORMAEMLSKI	840	CC	HCV antigens comprise one or more isolated antigens form a first region
Qy	537	QGLGLILRRHVGEGAVQWMMRLIAPASRGNHVSPTHYVPSRRRFAQALPVMWAPDYN	596	CC	of the HCV polyprotein, combining a biological sample with the solid
Db	841	QGLGLILRRHVGEGAVQWMMRLIAPASRGNHVSPTHYVPSRRRFAQALPVMWAPDYN	900	CC	support under conditions that allow HCV antibodies when present in the
Qy	597	PPLVETWKKPDYBPPVHVGRSSRRFAQALPVMWAPDYNPPLVETWKKPDYBPPVHGRKT	656	CC	biological sample, to bind to the one or more HCV antigens, adding to the
Db	901	PPLVETWKKPDYBPPVHVGRSSRRFAQALPVMWAPDYNPPLVETWKKPDYBPPVHGRKT	960	CC	solid support a detectably labelled HCV multiple epitope fusion antigen
Qy	657	KRNTNRRPDQVPPGGQIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRTWAOQGY	716	CC	(MEFA), where the labelled MEFA comprises at least one epitope from the
Db	961	KRNTNRRPDQVPPGGQIVG-----RRGP-----PIPKARRPEGRTWAOQGY	1003	CC	same region of the HCV polyprotein as the one or more isolated antigens,
Qy	717	PWPLYGNKDRRSTCKSGKPGYWPWPKTKRNTNRRPDQVPPGGQIVGGVYLLPRGPR	776	CC	where the MEFA binds to the bound HCV antibody, and detecting complexes
Db	1004	PWPLYGNKDRRSTCKSGKPGYWPWPKTKRNTNRRPDQVPPGGQIVG-----RRGP-1056		CC	formed between the HCV antibody and the MEFA, if any, as an
Qy	777	LGVLATRKTSPIPKARRPEGRTWAOQGYPWPLYGNKDRRSTCKSGKPGYWPW	829	CC	first region of the HCV polyprotein and the MEFA, if any, as an
Db	1057	-----PIPKARRPEGRTWAOQGYPWPLYGNKDRRSTCKSGKPGYWPW	1099	CC	indication of HCV infection in the biological sample. The method is
AC	ADL66809	standard; protein; 1099 AA.		CC	useful for detecting hepatitis C virus (HCV) infection in a biological
XX	ADL66809;			CC	sample. This sequence represents the MEFA 7.1 polypeptide used in the
XX	03-JUN-2004	(first entry)		XX	scope of the invention.
XX	HCV multiple epitope fusion antigen 7.1 (MEFA 7.1) polypeptide.			Qy	Sequence 1099 AA;
XX				Qy	Query Match
XX				Qy	Best Local Similarity 90.5%; Score 4032; DB 8; Length 1099;
XX				Qy	Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;
XX				Qy	1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLHGHVHFGDNTAGCTS 60
XX				Qy	1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLHGHVHFGDNTAGCTS 60
XX				Qy	61 AGHENPLSTR-----
XX				Qy	61 AGHENPLSTR-----
XX				Qy	72 -----GCNCSIYPGHITGHRMAWKLS 93
XX				Qy	121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIAQNLSGNCNCSIYPGHITGHRMAWKLS 180
XX				Qy	94 AARTTSGFVSLFAPGAKQKQETHVTGGAARTTSGLTSLFSPGASQNIQLITS-----145
XX				Qy	181 AARTTSGFVSLFAPGAKQKQETHVTGGAARTTSGLTSLFSPGASQNIQLIVDFIPVENLE 240
XX				Qy	146 -----TDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL 197
XX				Qy	241 TTWRSPTVTDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL 300
XX				Qy	198 GFGAYMSKAHGIDPNIRTTGSPITYSYTGKFLADGGCGSGAYDIICDECHSTD 257



```
Db 301 GFCAYMSKAHGDIPNTRTGTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTD 360
Qy 258 ATSILGIGTGLDQAEAGARLVLATATPGSVTVPHNIEEVALSTTGEIPYGAIP 317
Db 361 ATSILGIGTGLDQAEAGARLVLATATPGSVTVPHNIEEVALSTTGEIPYGAIP 420
Qy 318 EVTKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 377
Db 421 EVTKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 480
Qy 378 GYTGDSDSVLDNCTC----- 392
Db 481 GYTGDSDSVLDNCTC----- 392
Qy 393 ----- 392
Db 541 AGERPSGMFSDSVLCECYDAGCAWYELTAEITVRLRAYMNTPGLPVCDHLEFWEQVF 600
Qy 393 ----- 392
Db 601 TGLTHIDAHFLSTQKSGENLPYLVAQATVCARAQAPPSPQDWMKCLIRLKP TLHGPT 660
Qy 393 ----- 392
Db 661 PLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSAACSGKPAIIPDREVLRYREFDEME 720
Qy 417 CSOHLPIYIEOGMLAEQFKALGLSGRGPALVPDKEVLYQYDEMECSQAAPYIEQA 476
Db 721 CSOHLPIYIEOGMLAEQFKALGLSGRGPALVPDKEVLYQYDEMECSQAAPYIEQA 780
Qy 477 QVIAHQFKEKVLGLIDNDQVVVTPDKEILYEAFADEMEECASKAALIEEGORMAEMLSKI 536
Db 781 QVIAHQFKEKVLGLIDNDQVVVTPDKEILYEAFADEMEECASKAALIEEGORMAEMLSKI 840
Qy 537 QGLLGLILRRHVGEAGVQWMMRLIAFASRGNHVSPTHYVPSRSPRPAQALPVWARP 596
Db 841 QGLLGLILRRHVGEAGVQWMMRLIAFASRGNHVSPTHYVPSRSPRPAQALPVWARP 900
Qy 597 PPLVETWKKPDYPPVHVGSRSSRRFAQALPVWARPDPYNPPLVETWKKPDYPPVHVG 656
Db 901 PPLVETWKKPDYPPVHVGSRSSRRFAQALPVWARPDPYNPPLVETWKKPDYPPVHVG 960
Qy 657 KRNTNRRPQDVKPPGGQIYGVYLLPRRGPRLGLVLA TRKTSPIPKARRPEGR 716
Db 961 KRNTNRRPQDVKPPGGQIYGVYLLPRRGPRLGLVLA TRKTSPIPKARRPEGR 1003
Qy 717 PWPLYGNKDRRSTGSKWGKPGYPPWPRKTKNTNRRPQDVKPPGGQIYGVYLLPRRGP 776
Db 1004 PWPLYGNKDRRSTGSKWGKPGYPPWPRKTKNTNRRPQDVKPPGGQIYGVYLLPRRGP 1056
Qy 777 LGVLATRKTSPIPKARRPEGR 829
Db 1057 -----PIPKARRPEGR 1099

RESULT 7
AAW34481
ID AAW34481 standard; protein; 1021 AA.
AC AC
XX AAW34481;
DT 25-MAR-2003 (revised)
DT 16-MAR-1998 (first entry)
XX XX
DE HCV antigen combination pSOD/c200/core.
XX KW PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;
XX KW C domain; S domain; NS5; HCV polypeptide; anti-HCV antibody; detection;
XX KW NS4.
XX OS Hepatitis C virus.
OS Synthetic.
```

```
XX FH Key Location/Qualifiers
FT Misc-difference 1..902 /note= "linker"
FT FT Misc-difference 1..154 /note= "bSOD fragment"
FT FT Misc-difference 155..159 /note= "linker"
FT FT Misc-difference 160..899 /note= "c200 (amino acids 1192-1931 of HCV polypeptide)"
FT FT Misc-difference 903..1021 /note= "c22 (amino acids 2-120 of HCV polypeptide)"
XX XX US5683864-A.
PN XX 04-NOV-1997.
XX XX 07-JUL-1992; 92US-00910760.
XX PR 18-NOV-1987; 87US-00122714.
PR PR 30-DEC-1987; 87US-00139886.
PR PR 26-FEB-1988; 88US-00161072.
PR PR 06-MAY-1988; 88US-00191263.
PR PR 26-OCT-1988; 88US-00263584.
PR PR 14-NOV-1988; 88US-00271450.
PR PR 17-MAR-1989; 89US-00325338.
PR PR 21-APR-1989; 89US-00341334.
PR PR 18-MAY-1989; 89US-00353896.
PR PR 04-APR-1990; 90US-00504352.
XX XX (CHIR ) CHIRON CORP.
XX XX Kuo G, Houghton M, Choo Q;
XX WPI; 1997-548976/50.
XX N-PSDB; AAT99982.
XX PT Combination of three hepatitis C virus antigens - used for detection of
XX PT specific antibodies to diagnose infection.
XX PS Example 6; Col 59-68; 57pp; English.
XX CC This sequence represents a Hepatitis c virus (HCV) antigen combination of
XX CC the invention. The HCV antigen combination comprises an antigen (Ag1)
XX CC comprising the C domain (i.e. amino acids (aa) 1-120 of the HCV
XX CC polypeptide), or its immunologically reactive fragment containing at
XX CC least 8 aa. It also comprises two additional antigens from two different
XX CC polypeptide domains, including at least 8 aa from the NS3, NS4, S or NS5
XX CC domains of the polypeptide, corresponding, respectively, to aa 1050-1640;
XX CC 1640-2000; 120-400 and 2000-3011 of the HCV polypeptide. Alternatively,
XX CC Ag1 contains at least 8 aa from the 1-122 or 9-177 aa regions of the HCV
XX CC polypeptide. These antigen combinations are used diagnostically to detect
XX CC anti-HCV antibodies, using any standard immunoassay format. These antigen
XX CC combinations have a broader range of reactivity with antibodies than any
XX CC antigen individually. (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 1021 AA;
```

```
Query Match 49.9%; Score 2222; DB 2; Length 1021;
Best Local Similarity 46.2%; Pred. No. 4.1e-150;
Matches 512; Conservative 35; Mismatches 111; Indels 450; Gaps 22;
Qy 1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60
Db 1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60
Qy 61 AGPHFNPLSTRGNCISYIPGHITGHRMAWKLSAARTSG-----FVSL----- 104
Db 61 AGPHFNPLSRK-----HGGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCII 114
Qy 105 -----FAPGAKQNETHTVGTGAARATSGTSLTSLFSPGASQNIQ----- 141
```

Db	115	GRTLVVHEKADDLGKGGNEESTK-TGNAGSRLACGVI-----GIAQNLEFGAVDFIPVEN	168
Qy	142	LITS-----TDNSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	195
Db	169	LETTMRSFPVFTDSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	228
Qy	196	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHS	255
Db	229	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHS	288
Qy	256	TDATSLTGTVLDOAETAGARLVLATATPGSVTVPHNPNEEVALSTTGEIPFYGKAI	315
Db	289	TDATSLTGTVLDOAETAGARLVLATATPGSVTVPHNPNEEVALSTTGEIPFYGKAI	348
Qy	316	PLEVIKGGRRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDAL	375
Db	349	PLEVIKGGRRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDAL	408
Qy	376	MTGYTGDFPSVDCNTC-----	392
Db	409	MTGYTGDFPSVDCNTCVTQTVDPSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYR	468
Qy	393	-----	392
Db	469	FVAPGRPSGMFDSVLCYCDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWEG	528
Qy	393	-----	392
Db	529	VFTGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAAPPSPWDQWKCLIRLKPRTLHG	588
Qy	393	-----	392
Db	589	PTPLLYRLGAVQNEITLTHPTVKYIMTCMSADLEVTSTWVLVGGVLAALAAVCLSTGCV	648
Qy	393	-----ACSGKPAIIPDREVLVREFDEMESCQHLPIYEQGMLAEQFKALGL-----S	442
Db	649	VIVGRVLSGKPAIIPDREVLVREFDEMESCQHLPIYEQGMLAEQFKALGLLOTAS	708
Qy	443	RGKPAIIPDREVLVREFDEMESCQHLPIYEQGMLAEQFKALGLDNDQVV	497
Db	709	RQAE-VIAPAVQTNWQKLETFWAKHMWNFISGIQYLAGLSTLPG--NPALASLMAFTAAV	765
Qy	498	VTP-----DKEILYE-----AFDEMEECASKAALIEBQGMKSKIQ	537
Db	766	TSPLTTSTQTLFNILGNGWAAQAAPGAATAFVGAGLAGAAGSVGLGKVLIDILAGYGA	825
Qy	538	GLLG-----ILRRHVGRFEGAVQWMN	558
Db	826	GVAGALVAFKIMGEVPSDEDLVNLLPAIILSPGALVGVVCAALLRRHVGRFEGAVQWMN	885
Qy	559	RLIAFASRGNHVSPTTHYVPSRSRFAQALPWARPDYNPPLVETWKKPDYEPVPHGRSS	618
Db	886	RLIAFASRGNHVS-----GNSS	903
Qy	619	RRFAQALPWARPDYNPPLVETWKKPDYEPVPHGRKTRNRRPQDVKPPGCGQIVGG	678
Db	904	T-----NP-----KPKQ-----KKKRNTRNRPQDVKPPGCGQIVGG	934
Qy	724	VYLLPRGPRGLVATRKTS-----PIPKARPEGRWQAGVPMPLYGNK-----	724
Db	935	VYLLPRGPRGLVATRKTSERSQPRRRQPIPKARPEGRWQAGVPMPLYGNCGCW	994
Qy	725	-----DRRTSKSGWKPKFPGMPKTRKN	747
Db	995	AGWLLSPRGSRPSKGPDT---PRRRSRN	1019
RESULT 8			
AAW40039			
ID AAW40039 standard; protein; 1021 AA.			
XX			
AC AAW40039;			
XX			

DT	26-MAY-1998	(first entry)
XX	Fusion protein c200/c22.	
DE	Hepatitis C virus C domain; HCV; immunological activity; c200/c22;	
XX	NS3 domain; NS4 domain; S domain; NS5 domain; fusion protein.	
KW	Synthetic.	
XX	Hepatitis virus.	
OS	US5712087-A.	
XX	27-JAN-1998.	
XX	12-MAY-1995; 95US-00440519.	
XX	04-APR-1990; 90US-00504352.	
PR	07-JUL-1992; 92US-00910760.	
XX	(CHIR ) CHIRON CORP.	
XX	Kuo G, Houghton M, Choo Q;	
PI	WPI; 1998-119973/11.	
XX	N-PSDB; AAV09990.	
DR	Immunosays for hepatitis C virus antibodies - using combinations of	
DR	antigenic fragments of HCV polyprotein.	
XX	Example 6; Fig 4; 59pp; English.	
XX	This sequence represents a fusion protein constructed from the hepatitis	
CC	C virus core domain (which is situated at the carboxy terminus of the	
CC	fusion protein) and a c200 construct (a fusion of the NS3 and NS3	
CC	domains). This protein used in the construction of novel combinations of	
CC	HCV antigens that have a broader range of immunological activity than any	
CC	single HCV antigen. An example of such an antigen given in this	
CC	specification comprises a first antigen containing at least 8 amino acids	
CC	of the C domain of the HCV polyprotein and a second antigen comprising at	
CC	least 8 amino acids of the NS3 domain, the NS4 domain, the S domain or	
CC	the NS5 domain of the HCV polyprotein in the form of a fusion protein, a	
CC	physical mixture or bound to a solid matrix	
XX	Sequence 1021 AA;	
SQ	Query Match 49.9%; Score 2222; DB 2; Length 1021;	
	Best Local Similarity 46.2%; Pred. No. 4.1e-150;	
	Matches 512; Conservative 35; Mismatches 111; Indels 450; Gaps 22;	
QY	1 MATKAVCVLKGDPVOGIIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60	
DB	1 MATKAVCVLKGDPVOGIIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60	
QY	61 AGPHFNPLSTRGNCSTIYPGHITGHRMAWKLSAARTTSG-----FVSL----- 104	
DB	61 AGPHFNPLSRK-----HGGPKDEERHVGDLGNVTADKDGADVSDSIEDSVISLGDHCII 114	
QY	105 -----FAPGAKQNEHTVTGAAARTTSGLTSLFSPGASQNIQ----- 141	
DB	115 GRTLWVHEKADDLGKGGNEESTK-TGNAGSRLACGVI-----GIAQNLEFGAVDFIPVEN 168	
QY	142 LITS-----TDNSSPPVVPQSFQVAHLHAPTSGGKSTKVPAAAYAAQGYKVLVLPNSVAA 195	
DB	169 LETTMRSPVFTDSSPPVVPQSFQVAHLHAPTSGGKSTKVPAAAYAAQGYKVLVLPNSVAA 228	
QY	196 TLGFGAYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHS 255	
DB	229 TLGFGAYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHS 288	
QY	256 TDATSLIGTGLDQAEATAGARLWLATATPPGSVTVPHNPNEEVALSTTGEIPFYGKAI 315	
DB	289 TDATSLIGTGLDQAEATAGARLWLATATPPGSVTVPHNPNEEVALSTTGEIPFYGKAI 348	

```
Qy 316 PLEVIKGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSIVITSGDVVVVATDAL 375
Db 349 PLEVIKGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSIVITSGDVVVVATDAL 408
Qy 376 MTGYTGDFDSVIDCNTC----- 392
Db 409 MTGYTGDFDSVIDCNTCVTQTVDPSLPTFTIETITLPQDAVSRTQRRGTRGKPGIYR 468
Qy 393 ----- 392
Db 469 FVAPGERPSGMFSSVLCYDAGCAWVELTPAETTVRLRAYMTNPGLPVQCQDHLFWEQ 528
Qy 393 ----- 392
Db 529 VFTGLTHIDAHFLSQTKSGENLPYLWAYQATVCARAQAPPSPWDQMKCLIRLKLPTLHG 588
Qy 393 ----- 392
Db 589 PTPLLYRLGAVQNEITLTHPTVKYIMTCMSADLEVVTSTWVLVGVLAALAAAYCLSTGCV 648
Qy 393 -----ACSGKPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----S 442
Db 649 VIVGRVLSGKPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGLQTAS 708
Qy 443 RGGKPAIVDPKEVLYQQVD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLLDNDQVV 497
Db 709 RQAE-VIAPAVQTNWQKLETFWAKHMMNFISIQIYLAGLSLTPG--NPATIASLMAFTAAY 765
Qy 498 VTP--DKEILYE-----AFDEMEECASKAALIEEGORMAELKSKIQ 537
Db 766 TSPUTTSQTLNFGWAAQLAAPGAATFVAGLAGAAGSGLGKVLIDILAGYGA 825
Qy 538 GLLG-----ILRRHVGFGEGAVQWMN 558
Db 826 GVAGALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGFGEGAVQWMN 885
Qy 559 RLIAFASRGNHVSPTHYVPSRRRFAQALPWARPDYNPLVETWKKPDYPPVPHGRSS 618
Db 886 RLIAFASRGNHVS-----GNSS 903
Qy 619 RRFAQALPVWARPDYNPLVETWKKPDYPPVPHGRKTKRNTNRRPQDVKPPGGQIVGG 678
Db 904 T-----NP-----KPO-----KKNKNTNRRPQDVKPPGGQIVGG 934
Qy 679 VYLLPRRGPRLGLVLRKTS-----PIPKARRPEGRRTWAQPGYPWPLYGNK----- 724
Db 935 VYLLPRRGPRLGVLRATKTSERSQPRGRQPIPKARRPEGRRTWAQPGYPWPLYNCGCGW 994
Qy 725 -----DRRSTGKSWGKPGYPWPRKTKRN 747
Db 995 AGWLLSPRGRSPSWGPTD---PRRRSRN 1019

RESULT 9
AAE22050
ID AAE22050 standard; protein; 1021 AA.
AC
XX
XX AAE22050;
XX
DT 16-JUL-2002 (first entry)
XX
DE PSOD/c200/core expression plasmid protein.
XX
XX Hepatitis C virus; HCV; antigen; C domain; polyprotein; NS3 domain;
KW NS4 domain; S domain; NS5 domain; PSOD/c200/core plasmid.
XX
OS Hepatitis C virus.
OS Unidentified.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT 1..154
FT /note= "hsOD"
```

```
Region 155..159
/Note= "Linker region"
Region 160..899
/Note= "HCV c200"
Region 900..902
/Note= "Linker region"
Region 903..1021
/Note= "HCV c22"
US6312889-B1.
06-NOV-2001.
12-MAY-1995; 95US-00440549.
04-APR-1990; 90US-00504352.
07-JUL-1992; 92US-00910760.
(CHIR ) CHIRON CORP.
Houghton M, Choo Q, Kuo G;
WPI; 2002-040268/05.
N-PSDB; AAD35044.
Combination of hepatitis C viral (HCV) antigens, useful in improved
immunocassay for detecting HCV antibodies.
Example 6; Fig 4; 58pp; English.
The invention relates to combination of hepatitis C viral (HCV) antigens
that have a broader range of immunological reactivity than any single HCV
antigen. The combinations consist of an antigen from the C domain of the
HCV polyprotein, and at least one additional HCV antigen from either the
NS3 domain, the NS4 domain, the S domain, or the NS5 domain and are in
the form of fusion protein, a simple physical mixture, or the individual
antigens commonly bound to a solid matrix. The combinations of antigens
provides broad range immunoassays for anti-HCV antibodies. The invention
therefore provides a method for detecting antibodies to HCV in a mammal
suspected of containing such antibodies. The present sequence is a
CC protein encoded by pSOD/c200/core expression plasmid DNA containing HCV
CC coding sequence
CC Sequence 1021 AA;
SQ
Query Match 49.9%; Score 2222; DB 5; Length 1021;
Best Local Similarity 46.2%; Pred. No. 4.1e-150;
Matches 512; Conservative 35; Mismatches 111; Indels 450; Gaps 22;
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60
Qy 61 AGPHFNPLSRK-----HGGPKDBERHVGDLGNVTADKGVADVSDVIESVLSGDHCII 114
Db 61 AGPHFNPLSRK-----HGGPKDBERHVGDLGNVTADKGVADVSDVIESVLSGDHCII 114
Qy 105 -----FAPGAKQNETHTVTGGAARTTSLTSLFSPGASQNIQ----- 141
Db 115 GRTLNVHEKADDLCKGGNEESTK-TGNAGSLACGVI-----GIAQNLEFGAVDFIPVEN 168
Qy 142 LITS-----TDNSSPPVPOSFOVAHLHAPTGGSKSTKVPAAAYAAQYKVLVLPNSVAA 195
Db 169 LETTMRSPVFTDNSSPPVPOSFOVAHLHAPTGGSKSTKVPAAAYAAQYKVLVLPNSVAA 228
Qy 196 TLGFGAYMSKAHGDIPNIRTVRTITTSPTTYTYGKFLADGGCGGAYDIICDECHS 255
Db 229 TLGFGAYMSKAHGDIPNIRTVRTITTSPTTYTYGKFLADGGCGGAYDIICDECHS 288
Qy 256 TDATSIILGIGTVLDOAETAGARLVVLTATATPPGVTVPHPNIEEVALSTTGEIPFYGKAI 315
Db 289 TDATSIILGIGTVLDOAETAGARLVVLTATATPPGVTVPHPNIEEVALSTTGEIPFYGKAI 348
```

QY	316	PLEVIKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	375
Db	349	PLEVIKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	408
QY	376	MTGYTGDFSVIDCNC	392
Db	409	MTGYTGDFSVIDCNCVTQTVDFTFTIETITLPQDAVSRTQRRGTRGKPGIYR	468
QY	393	-----	392
Db	469	FVAPGRPSGMFSSVLCYDAGCAWYELTPAETTVRLRAYMNTGLPVCQDHLFEWEG	528
QY	393	-----	392
Db	529	VFTGLTHIDAHFLSQTKSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKP TLHG	588
QY	393	-----	392
Db	589	PTPLLYRLGAVQNEITLTPHTVKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCV	648
QY	393	-----ACSGKPAIIPDREVLYREFDEMECSOHLPIYIEQGMMLAQFKOKALGL---	442
Db	649	VIVGRVLSGKPAIIPDREVLYREFDEMECSOHLPIYIEQGMMLAQFKOKALGLLOTAS	708
QY	443	RGKPAIVDPKVELYQOYD-----EMECSSQAAPYIEQAQVIAHOFKKEKVLGLINDQVV	497
Db	709	QRAE-VIAPAVQTNQKLETFWAKHWNFIISGQYLAGLSTLPG--NPATASLMAFTAAV	765
QY	498	VTP-----DKEILYE-----AFDEMECSAKAALIEBQRMAMELKSKIQ	537
Db	766	TSPLTTSQTLILNIGWVAALAAAPGAATAFVAGLAGAAGSVGLGLVILIDILAGYGA	825
QY	538	GLAG-----ILRRHVGPGGAVQWMN	558
Db	826	GVAGALVAFKIMGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGGAVQWMN	885
QY	559	RLIAFASRGNHVSPTHYVPSRRAQALPVMARPDYNPLVETWKKPDYEPVPHGRSS	618
Db	886	RLIAFASRGNHVS-----GNSS	903
QY	619	RRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRKTKNTNRRPQDVKPGGGQIVGG	678
Db	904	T-----NP-----KPO-----KKNKENTNRRPQDVKPGGGQIVGG	934
QY	679	VYLLPRGRLGLVLRKTS-----PIKARRPEGRWQAQGYPWPLYGNK----	724
Db	935	VYLLPRGRLGLVLRKTSERSQPRGRQPIPKARRPEGRWQAQGYPWPLYGNKCGW	994
QY	725	-----DRRSTGKSGKPGYEPWPKTKRN	747
Db	995	AGWLLSPRGRSPSWGPTD---PRRSRN	1019
RESULT 10			
AAP92041			
ID	AAP92041 standard; protein; 1766 AA.		
XX			
AC	AAP92041;		
XX			
DT	09-SEP-2004 (revised)		
DT	25-MAR-2003 (revised)		
DT	02-MAR-1990 (first entry)		
XX	Hepatitis C virus (HCV) protein of cDNA clone inserts.		
XX			
KW	Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).		
XX			
OS	Hepatitis C virus.		
OS	Unidentified.		
XX			
PN	EP318216-A.		
XX			
PD	31-MAY-1989.		
XX	18-NOV-1988;	88EP-00310922.	
XX	18-NOV-1987;	87US-00122714.	
PR	30-DEC-1987;	87US-00139886.	
PR	26-FEB-1988;	88US-00161072.	
PR	06-MAY-1988;	88US-00191263.	
PR	26-OCT-1988;	88US-00263584.	
PR	14-NOV-1988;	88US-00271450.	
XX	(CHIR ) CHIRON CORP.		
PA	(CHIR ) CHIRON CORP.		
PI	Houghton M, Choo QL, Kuo G;		
XX	WPI: 1989-159274/22.		
DR	N-PSDB; AAN92097.		
XX	Purified hepatitis C virus - and associated nucleic acids and polypeptide(s).		
PS	Claim 13; Fig 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.		
XX	It is the sequence encoded in the open reading frame of hepatitis C virus cDNA inserts in clones 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could be used in immunosay reagents and vaccines and to generate antibodies useful in diagnosis and passive immunotherapy for HCV infection/non-A, non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PI field.)		
CC	Revised record issued on 09-SEP-2004 : Correction to DE line		
CC	Sequence 1766 AA;		
XX	Query Match	36.5%;	Score 1627.5; DB 1; Length 1766;
XX	Best Local Similarity	36.6%;	Pred. No. 3.7e-107;
XX	Matches	398;	Conservative 26; Mismatches 81; Indels 581; Gaps 15;
QY	146	TDNSSPPVPOSPQVAHLHAPTGSXSTKVPAAAYAAQGVKVLNPSVAATLFGAYMSK	205
Db	495	TDNSSPPVPOSPQVAHLHAPTGSXSTKVPAAAYAAQGVKVLNPSVAATLFGAYMSK	554
QY	206	AHGIDPNIRTVRTITGSPITVSTYVGKFLADGGCGGAYDIIICDECHSTDATSILGIG	265
Db	555	AHGIDPNIRTVRTITGSPITVSTYVGKFLADGGCGGAYDIIICDECHSTDATSILGIG	614
QY	266	TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYKKAIPLEVIKGRH	325
Db	615	TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYKKAIPLEVIKGRH	674
QY	326	LIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS	385
Db	675	LIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS	734
QY	386	VIDCNC	392
Db	735	VIDCNCVTQTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYFVAPGERPSG	794
QY	393	-----	392
Db	795	MPDSSVLCYDEGCWYELTPAETTVRLRAYMNTGLPVCQDHLFEWEGVFTGLTHIDA	854
QY	393	-----	392
Db	855	HFLSQTKSGENLPYLVAQATVWARAQPAPPPSDQMWKCLIRLKP TLHGTPPLLYRLGA	914
QY	393	-----ACSG	396
Db	915	VQNEITLTPHTVKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG	974
QY	397	KPAIIPDREVLYREFDEMECSOHLPIYIEQGMMLAQFKOKALGL---	452

```

Db 975 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGLLOTASQAE-VIAPA 1033
QY 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1034 VQTNWQKLETFWAKHWNFIISGIQYLAGSLTPG--NPAIASLMAFTAATVTSPLTTQTL 1091
QY 505 LYE-----AFDEMECASKAALIEGQRMMAEMLSKIOGLLG-----541
Db 1092 LFNILGGWVAQAALAPGAATAFVAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFK 1151
QY 542 -----TLRRHVGPGEAVQWMMRLIAFASRGN 568
Db 1152 IMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVQWMMRLIAFASRGN 1211
QY 569 HVSPTHYVPS-----578
Db 1212 HVSPTHYVPSDAAARVTAISSLTVTQLRLRLHQLWISSECTTPCSGSLRDIWDWICEV 1271
QY 579 -----578
Db 1272 LSDFKTLWAKMLPQLPGIPFVSCQGYGYVVDGIMHTRCHGCAEITGHVXNGTMRIV 1331
QY 579 -----578
Db 1332 GPRTCRNWMSGTFPINAYTTCPTLPAPNVTFFALMRVSAEYVEIRQVPSPEFTLDG 1391
QY 579 -RSRFA-----QALPWARPDY-----595
Db 1392 VRLHRFAPPCKPLREEVSFRVGLHEYPVGSQCEPEPDVAVLTSMITDPSHITAEAG 1451
QY 596 -----NPP-----LVT---WKK-----605
Db 1452 RRLARGSPPSVASSASQSLKATCTANHDSFDAELIEANLLWRQEMGNITRVESE 1511
QY 606 -----PDYEPVVG-----RSSRRAQALPVWAPDYNPPVETWKKPDYE 647
Db 1512 NKKVILSFDPLVAEEDEREISVPAELIRKSRRAQALPVWAPDYNPPVETWKKPDYE 1571
QY 648 PPVVG 653
Db 1572 PPVVG 1577

RESULT 11
ID AAP90164 standard; protein; 2261 AA.
XX AAP90164;
AC AAP90164;
DT 25-MAR-2003 (revised)
DT 01-NOV-1989 (first entry)
XX
DE Peptide encoded by composite hepatitis C virus cDNA.
XX
KW Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.
XX
OS Pan troglodytes.
XX
PN GB2212511-A.
XX
PD 26-JUL-1989.
XX
PF 18-NOV-1988; 88GB-00027024.
XX
PR 18-NOV-1987; 87US-00122714.
PR 30-DEC-1987; 87US-00139886.
PR 26-FEB-1988; 88US-00161072.
PR 26-OCT-1988; 88US-00263584.
XX
PA (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
XX

```

```

DR WPI; 1989-215054/30.
XX N-PSDB; AAN90331.
PT Hepatitis C virus gene - used for prodn. of polynucleotide probes
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment of
PT infection.
XX
PS Disclosure; Fig 32; 30pp; English.
XX
CC The sequence is the peptide encoded by the composite hepatitis C virus
CC (HCV) cDNA of AAN90331. The polypeptides are used to diagnose HCV-induced
CC NANBH, to raise antibodies for immunoassay or treatment, or to produce
CC vaccines. (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2261 AA;
Query Match 36.5%; Score 1624.5; DB 1; Length 2261;
Best Local Similarity 36.1%; Pred. No. 8.4e-107;
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;
QY 146 TNSSPVPVPSQFOVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db 586 TNSSPVPVPSQFOVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 645
QY 206 AHGIDPNIRTVRTITTTGSPITYSTYCKFLADGCGSGAYDIIICDECHSDTATSILGIG 265
Db 646 AHGIDPNIRTVRTITTTGSPITYSTYCKFLADGCGSGAYDIIICDECHSDTATSILGIG 705
QY 266 TVLDOAETAGARLVVLTATATPPGSVTPHPNIEVALSTTGEIPFYGKAIPLVYKGRH 325
Db 706 TVLDOAETAGARLVVLTATATPPGSVTPHPNIEVALSTTGEIPFYGKAIPLVYKGRH 765
QY 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGYTGDPS 385
Db 766 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGYTGDPS 825
QY 386 VIDNCTC-----392
Db 826 VIDNCTCVTQTVDFSLDPTFTIETILPQDAVSTQRRGRTGRGKPGIYRFVAFGERPSG 885
QY 393 -----392
Db 886 MFDSSVLCEYDAGCANVELTPAETTVRLRAYMNTPGLPVQODHLEFWGVFTGLTHIDA 945
QY 393 -----392
Db 946 HFLSOTKQSGENLPVLVAYQATVCARAQAPPSQWQMKLIRLKLPTLHGPTLLYRLGA 1005
QY 393 -----ACSG 396
Db 1006 VQNEITLTHPVTKYIMTMSADLEVTVTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG 1065
QY 397 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIYVD 452
Db 1066 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGLLOTASQAE-VIAPA 1124
QY 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1125 VQTNWQKLETFWAKHWNFIISGIQYLAGSLTPG--NPAIASLMAFTAATVTSPLTTQTL 1182
QY 505 LYE-----AFDEMECASKAALIEGQRMMAEMLSKIOGLLG-----541
Db 1183 LFNILGGWVAQAALAPGAATAFVAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFK 1242
QY 542 -----TLRRHVGPGEAVQWMMRLIAFASRGN 568
Db 1243 IMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVQWMMRLIAFASRGN 1302
QY 569 HVSPTHYVPS-----578
Db 1303 HVSPTHYVPSDAAARVTAISSLTVTQLRLRLHQLWISSECTTPCSGSLRDIWDWICEV 1362
QY 579 -----578

```

Db 1363 LSDFKTLKAKLMPQLPGIPFVSCQRGYKGVWRVGDGIMHTRCHGAEITGHVKNGTMRIV 1422  
QY 579 ----- 578  
Db 1423 GPRTRCNMWSGTFPINAYTTGPTCLPAPNYTFALMRVSAEYVEIRQVGFHYVTGMTT 1482  
QY 579 -----RSRFA-----QALPVMWARP 594  
Db 1483 DNLKPCQVPSEFFTELDGRLHRFAPPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPD 1542  
QY 595 Y-----NPP-----LV 600  
Db 1543 VAVLTSMLTDPSSHITAEAGRRRLARGSPPVASSASQLSAPSIKATCTANHDSPOAELI 1602  
QY 601 ET--WK-----PDYPPVVG-----RSRFAQALPV 627  
Db 1603 EANLLWRQMGNNITRVESENKVVILDSFDPLVAEEDEREISVPAETILRKSRFPAQALPV 1662  
QY 628 WARPDPNPPLVETWKPDYPPVVG 653  
Db 1663 WARPDPNPPLVETWKPDYPPVVG 1688

RESULT 12

AAP92050  
ID AAP92050 standard; protein; 2436 AA.

AC AAP92050;  
XX  
DT 09-SEP-2004 (revised)  
DT 25-MAR-2003 (revised)  
DT 02-MAR-1990 (first entry)

XX HCV protein of the cDNA inserts in clones K9-1 through 15e.

XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).

XX Hepatitis C virus.  
OS Unidentified.

XX EP318216-A.  
XX 31-MAY-1989.  
XX 18-NOV-1988; 88EP-00310922.  
XX 18-NOV-1987; 87US-00122714.  
XX 30-DEC-1987; 87US-00139886.  
XX 26-FEB-1988; 88US-00161072.  
XX 06-MAY-1988; 88US-00191263.  
XX 26-OCT-1988; 88US-00263584.  
XX 14-NOV-1988; 88US-00271450.

XX (CHIR ) CHIRON CORP.  
PA (CHIR ) CHIRON CORP.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-159274/22.  
DR N-PSDB; AAN92106.

XX Purified hepatitis C virus - and associated nucleic acids and  
PT polypeptides).

XX Claim 13; Fig 47-1-47-8; 139pp; English.

XX It is the sequence encoded in the open reading frame of hepatitis C virus  
CC (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could  
CC be used in immunoassay reagents and vaccines and to generate antibodies  
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,  
CC non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated  
CC on 25-MAR-2003 to correct PI field.)

CC Revised record issued on 09-SEP-2004 : Correction to DE line  
XX  
SQ Sequence 2436 AA;  
Query Match 36.5%; Score 1624.5; DB 1; Length 2436;  
Best Local Similarity 36.1%; Pred. No. 9.3e-107;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;  
QY 146 TONSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGVKVLNPSVAATLGFAYMSK 205  
Db 761 TONSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGVKVLNPSVAATLGFAYMSK 820  
QY 206 AHGIDNIRITGVTITGSPITYTYGKFLADGCGSGGAYDIIICDECHSTDATSILGIG 265  
Db 821 AHGIDNIRITGVTITGSPITYTYGKFLADGCGSGGAYDIIICDECHSTDATSILGIG 880  
QY 266 TVLDQAEATAGARLVLATATPPGSVTVPHNPIEVALSTTGEIPFYGKAIPLEVIKGRH 325  
Db 881 TVLDQAEATAGARLVLATATPPGSVTVPHNPIEVALSTTGEIPFYGKAIPLEVIKGRH 940  
QY 326 LIFCHSKKCDLAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMTGYTGFDS 385  
Db 941 LIFCHSKKCDLAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMTGYTGFDS 1000  
QY 386 VIDCNTC----- 392  
Db 1001 VIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPGERPSG 1060  
QY 393 ----- 392  
Db 1061 MFDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFEGVFTGLTHIDA 1120  
QY 393 ----- 392  
Db 1121 HFLSQTQSGENLPYLVAQVATVCARAQAPPPSDQWKNCLIRLKTPLHPTLLYRLGA 1180  
QY 393 -----ACSG 396  
Db 1181 VQNEITLTHPTKVIMTCHSADLEVTSTWLVGGVLAALAAVCLSTGCVVIVGRVVLSG 1240  
QY 397 KPAPIIDREVLYREFDEMEBECQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452  
Db 1241 KPAPIIDREVLYREFDEMEBECQHLPIYIEQGMMLAEQFKQKALGLLQATASQAB-VIAPA 1299  
QY 453 KEVLYQQYD-----EMEBCSQAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEI 504  
Db 1300 VQTNWQKLETFWAKHWNFIISGQYLAGLSTLPG--NPATASLMAFTAATVTSPLTTSQTL 1357  
QY 505 LYE-----AFDEMEBCASKAALIEBQORMAEMLSKIKOGLLG----- 541  
Db 1358 LFNILGWVAAQLAAPGAATAFVAGLAGAAGAAIGSVGLKVLIDILAGYGAGVAGALVAFK 1417  
QY 542 -----ILRRHVGPGECAVQWMMRLIAFASRGN 568  
Db 1418 IMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGECAVQWMMRLIAFASRGN 1477  
QY 569 HVSPTHYVRS----- 578  
Db 1478 HVSPTHYVPSDAAARVAILSSLTVTQLRLRLHQMISSECTTTCSSGSLRDIWDWICEV 1537  
QY 579 ----- 578  
Db 1538 LSDFKTLKAKLMPQLPGIPFVSCQRGYKGVWRVGDGIMHTRCHGAEITGHVKNGTMRIV 1597  
QY 579 ----- 578  
Db 1598 GPRTRCNMWSGTFPINAYTTGPTCLPAPNYTFALMRVSAEYVEVRQVGFHYVAGMIT 1657  
QY 579 -----RSRFA-----QALPVMWARP 594  
Db 1658 DNLKPCQVPSEFFTELDGRLHRFAPPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPD 1717

Qy	595 Y-----NPP-----LV 600
Db	1718 VAVLTSMLTDPSSHITAEAGRLRAGSPSVASSASQLSAPSLKATCTANHDSPPDAELI 1777
Qy	601 ET---WKK-----PDYEPVVHG-----RSSRRFAQALPV 627
Db	1778 EANLLWQEGNGNITRVESENKVVILDSFPLVAEEDEREISVPAEILRKRRRFAQALPV 1837
Qy	628 WARDYNPPLVETWKKPDYPPVVHG 653
Db	1838 WARDYNPPLVETWKKPDYPPVVHG 1863
RESULT 13	
AA	AAP90288
ID	AAP90288 standard; protein; 2436 AA.
AC	AAP90288;
XX	
DT	25-MAR-2003 (revised)
DT	19-JUL-2001 (revised)
DT	01-NOV-1989 (first entry)
XX	
DE	Peptide encoded by composite hepatitis C cDNA.
XX	
XX	Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
KW	
XX	
OS	Pan troglodytes.
XX	
FN	GB2212511-A.
XX	
PD	26-JUL-1989.
XX	
XX	18-NOV-1988; 88GB-00027024.
PF	
XX	
PR	18-NOV-1987; 87US-00122714.
PR	30-DEC-1987; 87US-00139886.
PR	26-FEB-1988; 88US-00161072.
PR	26-OCT-1988; 88US-00263584.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
PI	Houghton M, Choo QL, Kuo G;
XX	
XX	WPI; 1989-215054/30.
DR	N-PSDB; AAN90336.
XX	
PT	Hepatitis C virus gene - used for prodn. of polynucleotide probes
PT	polypeptide(s) and antibodies for diagnosis, prevention and treatment of
PT	infection.
XX	
PS	Disclosure; Fig 47-1 to 47-8; 30pp; English.
XX	
CC	The sequence is the peptide encoded by the composite hepatitis C virus
CC	(HCV) cDNA of AAN90336. The polypeptides are used to diagnose HCV-induced
CC	NANBH, to raise antibodies for immunoassay or treatment, or to produce
CC	vaccines. (N.B. This record was resubmitted to correct errors in the
CC	sequence.) (Updated on 25-MAR-2003 to correct PR field.)
XX	
XX	Sequence 2436 AA;
Query Match 36.5%; Score 1624.5; DB 1; Length 2436;	
Best Local Similarity 36.1%; Pred. No. 9,3e-107;	
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;	
Qy	146 TDNSSPPVPSQFVAHLHAPTGSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db	761 TDNSSPPVPSQFVAHLHAPTGSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 820
Qy	206 AHGIDPNRTGVRTTTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATSIIGIG 265
Db	821 AHGIDPNRTGVRTTTCSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATSIIGIG 880

RESULT 14  
AAB18540  
ID AAB18540 standard; protein; 2772 AA.  
XX

Qy	266 TVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 325
Db	881 TVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 940
Qy	326 LIFCHSKKKCDELAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS 385
Db	941 LIFCHSKKKCDELAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS 1000
Qy	386 VIDCNTC-----NPP-----LV 392
Db	1001 VIDCNTCTQTVDPSLDPTFTTITLPODAVSTQRRTGRGKPGIYRFVAPGERPSG 1060
Qy	393 -----
Db	1061 MFDSSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLTHIDA 1120
Qy	393 -----
Db	1121 HFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMKLIRLKLPTLHGPTLLYRLGA 1180
Qy	393 -----ACSG 396
Db	1181 VQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG 1240
Qy	397 KPATIPDREVLVREFDEMECSOHLPIYIEQGMMLAEQKOKALGL---SRGGKPAIVPD 452
Db	1241 KPATIPDREVLVREFDEMECSOHLPIYIEQGMMLAEQKOKALGLLQATASQAB-VIAPA 1299
Qy	453 KEVLYQQYD-----EMEECQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKXI 504
Db	1300 VQTNWQKLETFWAGHMNFISGIQYLAGLSLPG--NPATASLMAFTAATVSPLTTSQTL 1357
Qy	505 LYE-----AFDEMEECASKAALIEEGQMAEMLKSKIOGLG-----541
Db	1358 LFNILGGWAAQLAAPGAATAFVCGAGLAGAIGSVGLGKVLIDILAGYGAGVAGALVAFK 1417
Qy	542 -----TLRRHVGPGEAGVOMNRLIAFASRGN 568
Db	1418 IMSGEVPTEDLVNLLPALISPGALVGVVCAAILRRHVGPGEAGVOMNRLIAFASRGN 1477
Qy	569 HVSPTHYVPS-----578
Db	1478 HVSPTHYVPSDAAARVTAISSLVTQLLRLRHQWISSECTTPCSGSLWRDINDWICEV 1537
Qy	579 -----578
Db	1538 LSDFKTLWAKLMPQLPGIPFVSCQYGYVWRVDGIMHTRCHGAEITHGVKNGTMRIV 1597
Qy	579 -----578
Db	1598 GPTRCNMWSGTFPINAYTTGCTPLPAPNTYFALMRVSAEYVEIROGVDFHYVTGWT 1657
Qy	579 -----RSRFA-----QALPVWARP 594
Db	1658 DNLKPCQVPSPFEFTDGLVRLHRFAPPCKPLLRREEVSFRVGLHEYPVGSQLPCEPBD 1717
Qy	595 Y-----NPP-----LV 600
Db	1718 VAVLTSMLTDPSSHITAEAGRLRAGSPSVASSASQLSAPSLKATCTANHDSPPDAELI 1777
Qy	601 ET---WKK-----PDYEPVVHG-----RSSRRFAQALPV 627
Db	1778 EANLLWQEGNGNITRVESENKVVILDSFPLVAEEDEREISVPAEILRKRRRFAQALPV 1837
Qy	628 WARDYNPPLVETWKKPDYPPVVHG 653
Db	1838 WARDYNPPLVETWKKPDYPPVVHG 1863



AC AAB18540;  
XX  
DT 15-JAN-2001 (first entry)  
XX Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.  
DE  
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;  
KW viral infectivity; viral replication.  
XX  
XX Hepatitis C virus.  
XX  
XX EP1034785-A2.  
XX  
XX 13-SEP-2000.  
XX  
XX 16-MAR-1990; 2000EP-00109602.  
XX  
XX 17-MAR-1989; 89US-00325338.  
PR 20-APR-1989; 89US-00341334.  
PR 18-MAY-1989; 89US-00355002.  
PR 16-MAR-1990; 90EP-00302866.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Houghton M, Choo Q, Kuo G;  
XX  
XX WPI; 2000-566891/53.  
DR N-PSDB; AAV5296.  
XX  
XX Novel composition comprising a hepatitis C virus antisense polynucleotide  
PT which is complementary to or corresponds to a sense strand of the virus  
PT genome, and selectively hybridizes to it.  
XX  
XX Example; Fig 16; 75pp; English.  
XX  
XX The specification describes a pharmaceutical composition which comprises  
CC a hepatitis C virus (HCV) antisense polynucleotide. The HCV is  
CC characterized by a positive stranded RNA genome which has 40% homology at  
CC the polypeptide level to a HCV polypeptide. The antisense polynucleotide  
CC binds to cellular polynucleotides which enhance and/or are required for  
CC viral infectivity, replicative ability or chronicity. The antisense  
CC polynucleotides may also be designed to bind with high specificity, to be  
CC of increased stability, to be stable and to have low toxicity. The  
CC composition also comprises an agent which causes viral RNA to be  
CC inactive. The composition is used for preventing HCV replication in a  
CC system. The present sequence is encoded by a novel HCV cDNA sequence,  
CC which is used in the course of the invention  
XX  
XX Sequence 2772 AA;  
XX  
XX Query Match 36.5%; Score 1624.5; DB 3; Length 2772;  
Best Local Similarity 36.1%; Pred. No. 1.1e-106;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;  
XX  
XX 146 TDNSPPVPPQSFQVLAHLHPTGSGKTKVPAAYAAQGYKVLNPSVAATLGGAYMSK 205  
DB 1097 TDNSPPVPPQSFQVLAHLHPTGSGKTKVPAAYAAQGYKVLNPSVAATLGGAYMSK 1156  
XX  
XX 206 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 265  
DB 1157 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 1216  
XX  
XX 266 TVLDQAEATAGARLVVLTATPPGVTTPHPNIEVALSTTGEIPYKAIPLVKGGRH 325  
DB 1217 TVLDQAEATAGARLVVLTATPPGVTTPHPNIEVALSTTGEIPYKAIPLVKGGRH 1276  
XX  
XX 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGTGDFDS 385  
DB 1277 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGTGDFDS 1336  
XX  
XX 386 VIDCNTC----- 392  
DB 1337 VIDCNTCVTQVDFSLDPTFTIETITLPQDAVSRTORRGTGRGKIYRFPVAPGPPSG 1396  
XX

QY 393 ----- 392  
DB 1397 MPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLBEWGVFTGLTHIDA 1456  
QY 393 ----- 392  
DB 1457 HFLSOTKQSGENLPYLVAQATVCARAQAPPSWDQMCKLIRLKPTLHGPTPLLYRLGA 1516  
QY 393 -----ACSG 396  
DB 1517 VQNEITLTHPVTKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVIYGRVVLGS 1576  
QY 397 KPAPIIDREVLRYEFDEMEECESQHLPIYIQGMMLEAFQKQKALGL-----SRGKPAIVPD 452  
DB 1577 KPAPIIDREVLRYEFDEMEECESQHLPIYIQGMMLEAFQKQKALGLLQATASQAE-VIAPA 1635  
QY 453 KEVLYQOYD-----EMEECSQAAPYIEQAOVIAHQFKKVKVLGIDNDQVVVTP---DKEI 504  
DB 1636 VQTNWQKLETFWAKHMMNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTL 1693  
QY 505 LYE-----AFDEMEECASKAALIEEGORMAEMLKSKIQGLLG----- 541  
DB 1694 LFNILGWWAAQLAAPGAATAFVAGLAGAAGTSGVGLGKVLIDILAGYGAGVAGALVAFK 1753  
QY 542 -----ILRRHVGPGEVAGVQMMNRLIAFASRGN 568  
DB 1754 IMSGEVPTEDLVNLLPALISPCALVGVVCAAILRRHVGPGEVAGVQMMNRLIAFASRGN 1813  
QY 569 HVSPTHYVPS----- 578  
DB 1814 HVSPTHYVPSDAAARVTAILSSLTVQLRLRHQWISSECTTPCSGSLRDIWDWICEV 1873  
QY 579 ----- 578  
DB 1874 LSDFKTWLAKMLPOLPGIPFVSCQGYGVWRVVDGIMHTRCHGAEITGHVKNGTMRIV 1933  
QY 579 ----- 578  
DB 1934 GPTRCNMWSGTFPINAYTTGCTPLPAPNYTFALWRVSAEEVVEIRQVGDHVVGTMTT 1993  
QY 579 -----RSRRFA-----QALPVHARPD 594  
DB 1994 DNLKPCQVPSPFEFFTELDGVRHLRFAPPCPKLLREEVSFRVGLHVEYVPGSQLPCEPEPD 2053  
QY 595 Y-----NPP-----LV 600  
DB 2054 VAVLTSMLTDPSSHITABAGRRRLARGSPFSVASSASQSLSKATCTANHDSPDAL 2113  
QY 601 ET---WKK-----PDYEPVYVHG-----RSSRFPALPV 627  
DB 2114 EANLLWRQMGGNITRVESKNVVLDSFDPPLVAEEDEREISVPAETLRKSRFAQALPV 2173  
QY 628 WARPDPNPPILVETWKKPDYEPVYVHG 653  
DB 2174 WARPDPNPPILVETWKKPDYEPVYVHG 2199  
RESULT 15  
ADN35976  
ID ADN35976 standard; protein; 2772 AA.  
XX  
XX AC ADN35976;  
XX  
XX AC  
DT 17-JUN-2004 (first entry)  
XX  
XX HCV cDNA clone #1 protein.  
XX Antiviral; Vaccine; hepatitis C virus infection; HCV infection.  
XX  
XX Hepatitis C virus.  
XX  
XX EP1394255-A2.  
PN

```

XX 03-MAR-2004.
XX 16-MAR-1990; 2003EP-00016585.
XX 17-MAR-1989; 89US-00325338.
XX 20-APR-1989; 89US-00341334.
XX 18-MAY-1989; 89US-00350002.
XX 16-MAR-1990; 90EP-00302866.
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
XX WPI; 2004-193149/19.
XX N-PSDB; ADN35977.
XX
XX Novel purified hepatitis C virus polypeptide comprising epitope encoded
XX by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C
XX virus.
XX
XX Example 1; Fig 16; 79pp; English.
XX
XX The present invention relates to hepatitis C virus (HCV) proteins and
XX cDNA sequences. The sequences are useful in immunoassays for detecting
XX antibodies directed against HCV antigen; preparing host cells transformed
XX with a recombinant polynucleotide; screening antiviral agents and
XX determining the effect of antiviral agent in inhibiting viral replication
XX in cell culture system; and developing vaccine for treating HCV
XX infection.
XX
XX Sequence 2772 AA;
XX
XX Query Match 36.5%; Score 1624.5; DB 8; Length 2772;
XX Best Local Similarity 36.1%; Pred. No. 1.1e-106;
XX Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;
XX
XX 146 TDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAQGYKVLNPNVAATLFGAYMSK 205
XX 1097 TDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAQGYKVLNPNVAATLFGAYMSK 1156
XX
XX 206 AHGIDPNIRTVRTITGSPITYSTYKFLADGCGSGAYDIIICDCHSDTATSILIG 265
XX 1157 AHGIDPNIRTVRTITGSPITYSTYKFLADGCGSGAYDIIICDCHSDTATSILIG 1216
XX
XX 266 TVLDQAEATAGARLVVLTATAPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 325
XX 1217 TVLDQAEATAGARLVVLTATAPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 1276
XX
XX 326 LIFCHSKKCDLAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDS 385
XX 1277 LIFCHSKKCDLAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDS 1336
XX
XX 386 VIDCNTC----- 392
XX 1337 VIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRQRTGRGKPGIYRFVAPGERPSG 1396
XX
XX 393 ----- 392
XX
XX 1397 MFDSSVLCEYDACCANYELTPAETTVRLRAYMNTPLPVQCQDHLFEWGVFTGLTHIDA 1456
XX 393 ----- 392
XX
XX 1457 HFLSOTKQSGENLPYLVAQYQATVCARAQAPPPSNDQWKCLIRLKLPTLHGPTPLLYRLGA 1516
XX 393 -----ACSG 396
XX
XX 1517 VQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVVLSG 1576
XX
XX 397 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKOKALGL-----SRGGKPAIYVPD 452
XX 1577 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKOKALGLLQATASQAE-VIAPA 1635

```

```

Qy 453 KEVLYQOYD-----EMEESQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1636 VQTNWQKLETFWAKHWNFISGQIYLAGUSTLPG--NPAIASLMAFTAATAVTSPTTSTQL 1693
Qy 505 LYE-----AFDEMBECASKAALIBEGQRMAMLSKSKIOQLLG----- 541
Db 1694 LFNILGGWAAQLAAPCAATAFVGAGLAGAAGSVGLGKVLIDILAGYGAGVAGAVAFK 1753
Qy 542 -----ILRRHVGPGGGAOVQMMRLIAFASRGN 568
Db 1754 IMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGGGAOVQMMRLIAFASRGN 1813
Qy 569 HVSPTHVPS----- 578
Db 1814 HVSPTHVPSDAAARVAILSLTVTLRLRHQWISSSECTTPCSGSWLRIWDWICEV 1873
Qy 579 ----- 578
Db 1874 LSDFKTWLAKALPQLPGIPFVSCQRGYKGVWRVDGIMHTRCHCGAEITGHVKNGTMRIV 1933
Qy 579 ----- 578
Db 1934 GPRTCRNMSGTPPINAYTTGPTPLPAPNYTFALMRVSAEYVEIRQVGDHYYVTGTT 1993
Qy 579 -----RSRREA-----QALPVMARPD 594
Db 1994 DNLKCPQVPSBFFTELDGVRILHRFAPCKPLLRBEVSPRVGLHVEYVPGSQLPCEPEPD 2053
Qy 595 Y-----NPP-----LV 600
Db 2054 VAVLTSMLTDPSSHITAEAGRRLAGSPSPSVASSASQLSAPSLKATCTANHDSDAELI 2113
Qy 601 ET---WKK-----PDYEPVVVHG-----RSSRRFAQALPV 627
Db 2114 EAMLLMRQEMGNITRVSESENKVWILDSFDPVLAEEDEBISVPABILRKSRRAQALPV 2173
Qy 628 WARPDPNPLVETWKKPDYEPVVVHG 653
Db 2174 WARPDPNPLVETWKKPDYEPVVVHG 2199

```

Search completed: November 7, 2005, 20:10:17  
Job time : 117.776 secs

17

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:00:21 ; Search time 17.1255 Seconds  
(without alignments)  
4657.604 Million cell updates/sec

Title: US-10-658-782-4  
Perfect score: 4455  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRSTGKSWGKPGYWP 829

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1624.5	36.5	3011	1 GNMV3	genome polyprotein
2	1596	35.8	3011	1 S40770	genome polyprotein
3	1559.5	35.0	3011	1 GNMVCH	genome polyprotein
4	1534	34.4	3010	1 S18030	genome polyprotein
5	1532.5	34.4	3010	1 A45573	genome polyprotein
6	1523.5	34.2	3010	1 GNMVTC	genome polyprotein
7	1522.5	34.2	3010	1 GNMVTW	genome polyprotein
8	1519	34.1	3010	1 GNMVCJ	genome polyprotein
9	1429.5	32.1	3033	1 JQ1303	genome polyprotein
10	1422.5	31.9	3033	1 GNMVJ8	genome polyprotein
11	1337	30.0	3014	1 JCS620	genome polyprotein
12	1128.5	25.3	386	2 S68016	ATPase/RNA helicase
13	954.5	21.4	876	2 PC2219	polypeptide - hepa
14	860	19.3	216	2 S21337	genome polyprotein
15	608	13.6	3005	2 T08841	polyprotein - dour
16	586.5	13.2	2370	2 T08839	polyprotein - mara
17	409.5	9.2	492	2 PS0326	polyprotein - hepa
18	390	8.8	513	4 T43640	superoxide dismuta
19	387.5	8.7	179	4 T43640	superoxide dismuta
20	385	8.6	118	2 S41346	genome polyprotein
21	385	8.6	441	2 S12707	genome polyprotein
22	382.5	8.6	154	1 DSHUC7	superoxide dismuta
23	381	8.6	369	2 S21471	genome polyprotein
24	381	8.6	513	2 PC1284	genome polyprotein
25	380	8.5	178	2 PS0388	genome polyprotein
26	377	8.5	189	2 S32740	polyprotein - hepa
27	377	8.5	550	2 JH0711	genome polyprotein
28	377	8.5	782	2 S18031	genome polyprotein
29	377	8.5	787	2 PN0677	hypothetical prote

ALIGNMENTS

RESULT 1

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)  
N/Contains: capsid protein C, envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004  
C/Accession: A39166; PQ0403; PQ0404  
R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A/Title: Genetic organization and diversity of the hepatitis C virus.  
A/Reference number: A39166; MUID:91172826; PMID:1848704  
A/Accession: A39166  
A/Molecule type: mRNA  
A/Residues: 1-3011 <CHO>  
A/Cross-references: UNIPROT:P26664; GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
J.Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992  
A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e A/Reference number: PQ0393; MUID:92268871; PMID:1316939  
A/Accession: PQ0403  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CHA>  
A/Cross-references: DDBJ:D10128  
A/Experimental source: isolates E-b16  
A/Accession: PQ0404  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CH2>  
A/Experimental source: isolates E-b17  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F1-115/Product: capsid protein C #status predicted <CPC>  
F116-191/Product: envelope protein M #status predicted <EPM>  
F192-389/Product: major envelope protein E #status predicted <MEB>  
F390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F1007-1615/Product: hepatitis virus #status predicted <NS3>  
F1230-1237/Region: nucleotide-binding motif A (P-loop)  
F1312-1317/Region: nucleotide-binding motif B  
F1316-1319/Region: DEXH motif  
F1316-1862/Product: nonstructural protein NS4a #status predicted <NSA>  
F1863-2013/Product: nonstructural protein NS4b #status predicted <NSB>  
F2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22.22

Query Match 36.5%; Score 1624.5; DB 1; Length 3011;  
Best Local Similarity 36.1%; Pred. No. 1.9e-92;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;

QY 146 TNSSEPPVPSQFQVHLHAPGSGKSTKVPAAAGYKVLNPSVAATLGFQAYMSK 205  
|||||



Qy	393	-----	392
Db	1586	LVAYQATVCAQAAPPSPDQMWKCLIRLKLPTLHGPTLLYRLGAVQGEVTLTHPVTKYI	1645
Qy	393	-----	411
Db	1646	MTCSADLEVTSTWLVGVGLAALAYCLSTGCWIVGRVLSGRPAIIPDREVLYREF	1705
Qy	412	DEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKKAIPDPKEVLYQOYD-----E	462
Db	1706	DEMECSQHLPIYIEQGMMLAEQFKQKALGLLQATASRAE-VIAPTQVNTWQKLEAFNAKH	1764
Qy	463	MECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVTTP-----DKELIYE-----	507
Db	1765	MNFIISQIYLAGLSTLPG-----NPAIASLMAFTAATVSLTTSQTLLENILGGMVAAQLAA	1822
Qy	508	-----AFDEMECSKAALIEBEGQMAEMLKSKIQGLLG-----	541
Db	1823	PGAATAFVSGSLAGAVSGVGLGRVLVDILAGYGAGVALVAPKINSGLPSTEDLVNL	1882
Qy	542	-----ILRRHVGPGEVAVQVMMNRLIAFASRGNHVSPTHYVPS-----	578
Db	1883	LPAILSPGALVGVVCAAILRRHVGPGEVAVQVMMNRLIAFASRGNHVSPTHYVPSDAAA	1942
Qy	579	-----	578
Db	1943	RVTAILSSLVTQLRLRLHOWLSSESTTPSCGSMRLDIWDWICEVLSDFKTLKLMPH	2002
Qy	579	-----	578
Db	2003	LPGIPFVSCQHGKGVWRGDGMHTRCHGAEITHGVKNQGMWIRVIGPKTCRNMMSGTPI	2062
Qy	579	-----	578
Db	2063	NAYTGPCTPLPAPNYTALMRVSAEYVEIRRVGDHFHYVTGTTDNLKCPQVSPSEFF	2122
Qy	579	-----RSRRFA-----QALPVWARPDI-----	595
Db	2123	TELDGVLRRHTAPPCKPLLRBEVSPRVGLHDYVGSQLPCEPEPDVAVLTSLMTDPSHIT	2182
Qy	596	-----NPP-----LVET-----WKK-----	605
Db	2183	AAAGRLRARGSPSPSEASSASQASPSLKATCTTINHSDPAELIEANLLWRQENGGNIT	2242
Qy	606	-----PDYEPVVVHG-----RSSRFAQALPVWARPDPYNPPLIVETWK	642
Db	2243	RVESENKVVILDSFDPLVAEEDEREISVPAELIRKSRFTQALPIWARPDPYNPPLIVETWK	2302
Qy	643	KPDYEPVVVHG	653
Db	2303	KPNYEPVVVHG	2313
RESULT 3			
GNMVCH			
genome polyprotein - hepatitis C virus (strain H)			
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
A:Note: host Homo sapiens (man)			
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004			
R:Accession: A36814; A41546			
R:Inchouape, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M. submitted to GenBank, July 1992			
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C			
A:Reference number: A36814			
A:Molecule type: genomic RNA			
A:Residues: 1-3011 <INC>			
A:Cross-references: UNIPROT:P27958; GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738			
R:Inchouape, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M. Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991			
A>Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar			

A:Reference number: A41546; MUID:92052256; PMID:1658800			
A:Contents: annotation			
A:Note: neither amino acid nor nucleotide sequence is given			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura			
F:1-115/Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <BPM>			
F:192-389/Product: major envelope protein E #status predicted <MSE>			
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>			
F:1230-1237/Region: nucleotide-binding motif A (P-loop)			
F:1312-1317/Region: nucleotide-binding motif B			
F:1316-1319/Region: DEHX motif			
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>			
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>			
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 23			
Query Match 35.0%; Score 1559.5; DB 1; Length 3011;			
Best Local Similarity 34.9%; Pred. No. 2.2e-88;			
Matches 388; Conservative 32; Mismatches 80; Indels 611; Gaps 15;			
Qy	146	TDNSPPVPOSFOVAHLHAPTGSKSTKPAAYAAQYKVLNPNPSVAATLGFQAYMSK	205
Db	1211	TDNSPPAVPOSFOVAHLHAPTGSKSTKPAAYAAQYKVLNPNPSVAATLGFQAYMSK	1270
Qy	206	AHGIDPNIRTVRTITTTGSPITYSTYKFLADGGCGGAYDIICDECHSTDATSILGIG	265
Db	1271	AHGVDPIRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIICDECHSTDATSISGIG	1330
Qy	266	TVLQDAETAGARLVVLTATATPGSVTVPHNIEEVALSTTGEIPEYKAIPLEVKGGRH	325
Db	1331	TVLQDAETAGARLVVLTATATPGSVTVSHNIEEVALSTTGEIPEYKAIPLEVKGGRH	1390
Qy	326	LIFCHSKKKDELAAKVALGINAVYRGDLVSVIPTSGDVVVVATDALTMTGYTGDFDS	385
Db	1391	LIFCHSKKKDELAAKVALGINAVYRGDLVSVIPTSGDVVVVATDALTMTGYTGDFDS	1450
Qy	386	VIDCNC-----	392
Db	1451	VIDCNCVTQTVDPSLDPTFTTTLTPQDAVSRTRRGTGRGKGIYRVFVAPGERPSG	1510
Qy	393	-----	392
Db	1511	MPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHILGFWEVGTGLTHIDA	1570
Qy	393	-----	392
Db	1571	HFLSQTOSGENFPYLVAYQATVCARAQAPPPSWDMRKCLIRLKLHGTPLLYRLGA	1630
Qy	393	-----	ACSG 396
Db	1631	VQNEVTLTHPTIKYIMTQMSADLEVVSTWVLVGVGLAALAYCLSTGCWIVGRVLSG	1690
Qy	397	KPAIIPDREVLYREPDEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKKAIPVD	452
Db	1691	KPAIIPDREVLYREPDEMECSQHLPIYIEQGMMLAEQFKQKALGLLQATASRAE-VITPA	1749
Qy	453	KEVLYQOYDEMECSQAAPYIEQAQVIAHQFKEKVLGLINDQVV-----VT	499
Db	1750	VQTNWQKL-EVFWAKHMNFISGIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLT	1802
Qy	500	PKDEILYE-----AFDEMECSKAALIEBEGQMAEMLKSKIQGLLG- 541	
Db	1803	TQGTLLFNILGGVAAQLAAPGAATAFVGCAGLAGAALDSVGLKVLVDILAGYAGVAGA	1862
Qy	542	-----ILRRHVGPGEVAVQVMMNRLIAF	563
Db	1863	LVAFKIMSGEVPSSTEDLVNLLPAILSPGALAVGVVFAFILRRRRVGPGEVAVQVMMNRLIAF	1922
Qy	564	ASRGNHVSPTHYVPS-----	578

Db 1923 ASRGNHVSPTHYVPESDAAARVTAISSLITVQLRLHQMISSECTTPCSGSLWRDIWD 1982  
 QY 579 ----- 578  
 Db 1983 WICEVLSDFKTLWKALMPQLPGIPFVSCQGYRGVWRGDMITRCHGCAEITGHVNG 2042  
 QY 579 ----- 578  
 Db 2043 TMRIVGPRCTCKNMWGTFFINAYTTGPTPLPAPNYKFAKLRVSAEYVEIRRVGDFHYV 2102  
 QY 579 -----RSRRFA-----QALPV 589  
 Db 2103 SGMTTDLNLCPCQIPSPPEFFTELDGVLRLHFAPPCPKLLREEVSRFVGLHEYPVGSQLP 2162  
 QY 590 WARPDPY-----NRP----- 598  
 Db 2163 EPPDVAULTSMLTDPSSHITAAGRLARGLSPSPMASSASQSLAPSLKATCTANHDSP 2222  
 QY 599 ---LVET---WKK-----PDYEPVPHG-----RSSRRFA 622  
 Db 2223 DAELEALLWRQBMGGNITRVESENKVLWILDSFDPLVAEEDEREVSVPAILRKSRFA 2282  
 QY 623 QALPVWARPDPNPLVETWTKKPDYEPVPHG 653  
 Db 2283 PALPVWARPDPNPLVETWTKKPDYEPVPHG 2313

RESULT 4  
 S18030  
 genome polyprotein - hepatitis C virus (isolate JKI)  
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 A;Variety: isolate JKI  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
 C;Accession: S18030; S33570; A48332; S18029  
 R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
 A;Reference number: S18028  
 A;Accession: S18030  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3010 <HON>  
 A;Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
 A;Experimental source: isolate JKI from an individual  
 R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A;Reference number: A48332; MUID:93119270; PMID:8380322  
 A;Accession: S33570  
 A;Molecule type: genomic RNA  
 A;Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>  
 A;Cross-references: EMBL:X61591  
 A;Note: this sequence is inconsistent with the nucleotide translation  
 A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:P.121748)  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F;2-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-389/Product: major envelope protein E #status predicted <MEE>  
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>  
 F;1230-1237/Region: hepatitis C virus genome polyprotein  
 F;1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F;1316-1319/Region: DEXH motif  
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F;196, 209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate (As

Query Match 34.4%; Score 1534; DB 1; Length 3010;  
 Best Local Similarity 32.3%; Pred. No. 8.3e-87;  
 Matches 397; Conservative 41; Mismatches 115; Indels 676; Gaps 19;  
 QY 146 TDNSSPPVPSQVQVAHLHAPTSGSGSKTKVPAAYAAAGYKVLVNLNPSVAATLGFAYMSK 205  
 Db 1211 TDNSSPPAVPQTQVAHLHAPTSGSGSKTKVPAAYAAAGYKVLVNLNPSVAATLGFAYMSK 1270  
 QY 206 AHGIDNIRTVGTITTTGSPITYTYGKFLADCGSCGAYDIIICDCHSTDTATSLIGIG 265  
 Db 1271 AHGVNDINISTGVITTTGAPITYTYGKFLADCGSCGAYDIIICDCHSTDTATSLIGIG 1330  
 QY 266 TVLDQAEATAGARLVLAATATPPGCVTPPHENIEEVALSTTGEIPFYGAIPLEVIKGRH 325  
 Db 1331 TVLDQAEATAGARLVLAATATPPGCVTPPHENIEEVALPNTGEIPFYGAIPLEVIKGRH 1390  
 QY 326 LIFCHSKKCKDELAALKVNAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 385  
 Db 1391 LIFCHSKKCKDELAALKVNAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 1450  
 QY 386 VIDCNTC----- 392  
 Db 1451 VIDCNTCVTQTVDFSLDPTFTTITTTLPQDAVRSORRGRTGRGGIYRFVTPGERPSG 1510  
 QY 393 ----- 392  
 Db 1511 MPDSSVLCECYDAGCAWYELTPAVTSVRLRAYLNTPLPVCQVHLEPWESVFTGLTHIDA 1570  
 QY 393 ----- 392  
 Db 1571 HFLSQTQKAGENPYLVAYQATVCARAQAPPSPSQDQMKCLIRLKPTLHGPTLLYRLGA 1630  
 QY 393 -----AC-----SG 396  
 Db 1631 VQNEVTLTHPTIFKIMACHSADLEVTSVTLVGGVLAALAYCLTTGCVVIVGRITLSG 1690  
 QY 397 KPAAIPDREVLYREFDEMEECSQHLPIYIEQGMMLABQFKQKALGLSRGGKPAIVDPKEVL 456  
 Db 1691 RPAIIPDREVLYREFDEMEECSQHLPIYIEQGMMLABQFKQKALGLSRGGKPAIVDPKEVL 1740  
 QY 457 YQYDEMEECSQAAPYIE-----QAQVIAH-----QFKEVGLINDQV----- 497  
 Db 1741 -----KQAEAAAPVVESSKQWALEAFWAKHMNFISGIQYLAGLSTLPNPAIVSLMAF 1793  
 QY 498 -----VTPDKILY-----EADDEMEECSKAALIEEGORMAEMLK 533  
 Db 1794 TASITSPITTOHTLFLNLLGGWAAQLAPPSAASAFVAGIAGAAVSGISGLGVLDILA 1853  
 QY 534 SKIQGLLG-----ILRRHVGPGEAV 554  
 Db 1854 GYGAGVAGALVAFKMGSGEMPSTEDLVNLLPALSPCALVGVVCAAILRRHVGPGEAV 1913  
 QY 555 QMNRLLIAFASRGNHVSPTHYVP-----SSRRFAQALPWARDYN 596  
 Db 1914 QMNRLLIAFASRGNHVSPTHYVPDAAARVTKILSSLTITQRLRLHQ-----WINEDCS 1969  
 QY 597 PPLVETWKK----- 605  
 Db 1970 TPCSGSLWRDWDWICITVLTDFKTLQSKLLPRLPGDPFFSCQGVYRGVWGDVMTTC 2029  
 QY 606 -----PDY----- 608  
 Db 2030 PCGAQITGHVKGSMRIVGPKTCSNTWHTGTFPINAVTTGPTSPAPNYSRALWRVAEE 2089  
 QY 609 ----- 608  
 Db 2090 YVEVTRVGDPHYVTGMTDNDVKPCQVPAPEFFTEVDGVLHRYAPACKPLLRDEVTFQV 2149  
 QY 609 -----PPVV----- 613  
 Db 2150 GLNQFPVGSQPCPEPFDVTVLTSMLTDPSHITAEAKRLARGSPSSASSASQSLAP 2209  
 QY 614 -----HG----- 615



```

Db      2210 SLKATCTTRDSDPADLIEANLLWRQMGNIITRVESENKVWILDSFEPLRAEEDEREVS 2269
Qy      616 -----RSSRRFAQALPWARPDYNPLVETWKKPDYEPVFWHGRKTKRNTNRRPQDVKF 669
Db      2270 VAAEILKSRKFPALFIWAPSNFPLESKWDKPDYVPPVHGCPL-----PPTMAP 2322
Qy      670 PGGQIVGGVYLLPRRGR-----LGVLATRKTSPIPKARRPGRRTWAQPGYP 717
Db      2323 P-----IPPRKRKTWLTSTVSSALAEATKTFGSSGSSAVDSGTATAPDQP 2372
Qy      718 WPLYGNKDRSTGKSW-----CKPGYP 739
Db      2373 ---SDDGRGSDDESYSMPPEGEPCDP 2398

RESULT 5
A45573
genome polyprotein - hepatitis C virus (strain J7)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J7)
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1;
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein C; hydroxylase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <BPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus genome polyprotein
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match      34.4%; Score 1532.5; DB 1; Length 3010;
Best Local Similarity 33.5%; Pred. No. 1e-86;
Matches 373; Conserved 38; Mismatches 87; Indels 615; Gaps 13;

Qy      146 TDNSSPPVQSFQVHLHAPTSGSKTKVPAAYAAQYKVLNPNPSVAATLGFAYMSK 205
Db      1211 TDNSSPPAVPQTFQVHLHAPTSGSKTKVPAAYAAQYKVLNPNPSVAATLGFAYMSK 1270

Qy      206 AHGIDPNIRITGVRIITGSPITYSTYKFLADGCGSGAYDIIICDECHSDTATSIILGIG 265
Db      1271 AHGIDPNIRITGVRIITGSPITYSTYKFLADGCGSGAYDIIICDECHSDTATSIILGIG 1330

Qy      266 TVLQDAETAGARLVVLAATATPPGTVVPHNIEEVALSTTCIEIFYGKAIPLVFKGRH 325
Db      1331 TVLQDAETAGARLVVLAATATPPGTVVPHNIEEVALSTTCIEIFYGKAIPLVFKGRH 1390

Qy      326 LIFCHSKKKCDLAALKVALGINAVAYRGLDVSIVPTSGDGVVATDALTMTGTGDPDS 385
Db      1391 LIFCHSKKKCDLAALKSLGLGINAVAYRGLDVSIVPTSGDGVVATDALTMTGTGDPDS 1450

Qy      386 VIDNCNTC----- 392
Db      1451 VIDNCNTCTQTVDLSLDPFTTFTTTTPQDAVRSQRGRGTGRGRGIYRFTVPEPSPG 1510

```

```

Qy      393 ----- 392
Db      1511 MPDSSVLCEDYDAGCAWYELTPAETTVRLRAYLNTPLGVPQDHLFEWESVFTGLTHIDA 1570
Qy      393 ----- 392
Db      1571 HFLSOTKQAGDNFPYLVAQATVCARAQAPPPSHDQWKKLIRLKPTLHGPTPLLYRLGA 1630
Qy      393 -----AC-----SG 396
Db      1631 VQNEITLTHPTIKFIMACMSADLEVTWTWLVGGVLAALAAAYCLTTGTVVIGRIILSG 1690
Qy      397 KPAILIPDREVLYREFDEMESCQHLPIYIEQGMMAEQFKOKALGLSGRGRPAIVDPDEVL 456
Db      1691 RPAVVPDREVLYREFDEMESCQHLPIYIEQGMMAEQFKOKALGLSGRGRPAIVDPDEVL 1750
Qy      457 YQYDEMEE----- 465
Db      1751 ESRWRALAEAFWAKHMNFISGIQVLAGLSTLPGNPATASLMAPTASITSPLTNTQNTLLFN 1810
Qy      466 -----CSQAAPYIEQAQVIAHOFKEKVLGLINDQVVTTPDKEL-----YEAF 509
Db      1811 ILGQWVAQLAPPASAASAFVAGIAGAAIGSIGLVLV---DILAGYAGAGALVAF 1866
Qy      510 DEMEECASKA-----ALIEGQRMAMLMKSKIQGLLGLLRHRRHVGEGAVQMMRLI 561
Db      1867 KWMSGEAPSAEDLVNLLPAILSPCALVGVV-----CAAILRRHVGEGAVQMMRLI 1920
Qy      562 AFASRGNHVSPTHVP-----SRSRRFAQALPV-----WAPDYNPLVETWKK-- 605
Db      1921 AFASRGNHVSPTHVPESDAAARVTQILSLTTITQLKRLHQWINEDCSTPCSGSLWKDV 1980
Qy      606 ----- 605
Db      1981 WDMICTVLTDFKTWLSKLLPKLPGVFFSCQRYKGVWRGDGMQITTCPCGAQITGHVK 2040
Qy      606 -----PDY----- 608
Db      2041 NGSMRIVGPKTSNTWHTFTTPINAYTTGCTPSPAPNYSRALMRVAABEYVEITRVGDPH 2100
Qy      609 ----- 608
Db      2101 YVTGTTNDNVKPCQVQVAPBFFTELDGVLRLHRYAPACRPLLRBVTQVGLNQYLVGSOL 2160
Qy      609 -----EPPVW-----HG 615
Db      2161 PCEPDPVAVLTSLMTPDPSHITAEAKRLARGSPSSLASSASQLSAPSLKATCTTHD 2220
Qy      616 -----RSRR 620
Db      2221 SPDADLIEANLLWRQMGNIITRVESENKVWILDSFDPLRAEEDEREVSAAEILRKS 2280
Qy      621 FAQALPWARPDYNPLVETWKKPDYEPVFWHGRKTKRNTNRRPQDVKF 653
Db      2281 PPPALPIWARPDYNPLVETWKKPDYEPVFWHGRKTKRNTNRRPQDVKF 2313

RESULT 6
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J7)
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>

```

A/Cross-references: UNIPROT:P26663; EMBL:M58335; NID:g329770; PIDN:AA72945.1; PID:g32939

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F/2-115/Product: capsid protein C #status predicted <CPC>

F/116-191/Product: envelope protein M #status predicted <BPM>

F/192-389/Product: major envelope protein E #status predicted <MEE>

F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F/1007-1615/Product: hepatitis C virus #status predicted <NS3>

F/1230-1237/Region: nucleotide-binding motif A (P-loop)

F/1312-1317/Region: nucleotide-binding motif B

F/1316-1319/Region: DEXH motif

F/1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F/196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2222

	Query Match	34-2%	Score 1523.5	DB 1	Length 3010
	Best Local Similarity	33.2%	Pred. No. 3.8e-86		
	Matches 372	Conservative 39	Mismatches 80	Indels 629	Gaps 14
Qy	146	TDNSSPPVVFQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAATLFGAYMSK	205		
Db	1211	TDNSSPPVVFQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAATLFGAYMSK	1270		
Qy	206	AHGIDPNIRHTGRTITITGSPITVSTYCKFLADGCGSGGAYDIIICDECHSDTATSILGIG	265		
Db	1271	AHGIDPNIRHTGRTITITGSPITVSTYCKFLADGCGSGGAYDIIICDECHSDTATSILGIG	1330		
Qy	266	TVLDAQETAGARLVLATATPPGSGVTPHPNIEEVALSTTGEIPFYKAIPLKVIKGGRH	325		
Db	1331	TVLDAQETAGARLVLATATPPGSGVTPHPNIEEVALSTTGEIPFYKAIPLKVIKGGRH	1390		
Qy	326	LIFCHSKKCKDELAALKLVGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGYTGFDS	385		
Db	1391	LIFCHSKKCKDELAALKLVGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGYTGFDS	1450		
Qy	386	VIDCMTNC-----	392		
Db	1451	VIDCMTNCVTQTVDFSLDPTFTTETTVQDAVRSQRGRGRTGRRGTYRFTVTPGRPSG	1510		
Qy	393	-----	392		
Db	1511	MFDSVLCYCEYDAGCAYELTPTAETSVRLRAYLNTPLGVPQCQDHLFEWESVFTGLTHIDA	1570		
Qy	393	-----	392		
Db	1571	HFLSOTKQAGDNFPYLVAYQATVCARAQAPPPSDQMWKCLIRLKTPLHGTPLLYRLGA	1630		
Qy	393	-----AC-----	396		
Db	1631	VQNEVTLTHPTIKYIYACMSADLEVTSTWVLVGGVLAALAAAYCLTGTGSVVIVGRILLSG	1690		
Qy	397	KPAIIPDREVLVREFDEMEECSOHLPIYIEOGMMLAEQPKQKALGSRGKPAIVPDKVEL	456		
Db	1691	RPAIVPDRELLYQEFDEMEECSAHLPIYIEOGMMLAEQPKQKALGSRGKPAIVPDKVEL	1737		
Qy	457	YQYDEMEECSQAAPYIEQAQVIAHQPEK-VLGLINDQVV-----	497		
Db	1738	-----TATQKAAEAAAPVYESKWRALETTFWAKHWNFISIGIYLAGLSTLPGNPAIASLMAF	1793		
Qy	498	-----VTPDKELY-----	533		
Db	1794	TASITSPLTQTSLLFNLLGGVAAQLAPPSPAASAFVGCAGIAGAAGVSGIGLKVLDILA	1853		
Qy	534	SKIQGLG-----	554		
Db	1854	GYCAGVAGALVAFKWSGEMPMSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVAV	1913		
Qy	555	QMMNRLIAFASRGNHVSPTHTVPS-----	578		
Db	1914	QMMNRLIAFASRGNHVSPTHTVPSDAAARVTQILSSITITQLLXRLHWNEDCSTPCS	1973		

Qy	579	-----	578
Db	1974	GSWLRLVDWDMJCTVLTDFKTLWLSKLLPQLPGVPEFFSCQRYGKGVWRGDGIMQTTCPGA	2033
Qy	579	-----	578
Db	2034	QITGHVKNXGSMRIIVGPKTCSNTHGTPPINAYTTCTPCTSPAPNYSRALWRVAEEYEV	2093
Qy	579	-----	586
Db	2094	TRVGDFHYVTGTTTNDNVKPCQVPAPBEFFSEVDGVLRLHRYAPACRPLLRREEVTFQVGLNQ	2153
Qy	587	-----	598
Db	2154	YLVGSQLPCEPEPDVAULTSMLTDPDSHTAETAKRLARGSPPSLASSASQLSAPSLKA	2213
Qy	599	-----	613
Db	2214	TCTTHVSPDADLIEANLLRWQMGNGNITRVESENKVVVLDSDPLRABEDEREVSVPAE	2273
Qy	614	HGRSSRRPAQLPWARPDYNNPLVETWKKDDYEPVPHG	653
Db	2274	ILRKSKFPAAPIWARPDYNNPLLESWKDDYVPPVPHG	2313
RESULT 7			
GNWTV			
Genome polyprotein - hepatitis C virus (strain Taiwan)			
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
A:Note: host Homo sapiens (man)			
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004			
C:Accession: A40244			
F:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.			
Virology 188, 102-113, 1992			
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the			
A:Reference number: A40244; MUID:92230206; PMID:1314449			
A:Accession: A40244			
A:Molecule type: genomic RNA			
A:Residues: 1-3010 <HE>			
A:Cross-references: UNIPROT:P29846; GB:M84754			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural			
F:1-115/Product: capsid protein C #status predicted <CPC>			
F:116-161/Product: envelope protein M #status predicted <EPM>			
F:192-389/Product: major envelope protein E #status predicted <MEES>			
F:7390-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F:7390-729/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>			
F:1230-1237/Region: nucleotide-binding motif A (P-loop)			
F:1312-1317/Region: nucleotide-binding motif B			
F:1316-1319/Region: DEXH motif			
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>			
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>			
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077			
Query Match 34.2%; Score 1522.5; DB 1; Length 3010;			
Best Local Similarity 32.2%; Pred. No. 4,3e-86;			
Matches 397; Conservative 55; Mismatches 135; Indels 647; Gaps 21;			
Qy	41	EGLGHGFHYHVEFGDNTAGCTSGAPHENPLSTRGCNSIYP----	GHITGHRMAWKLGAAR 96
Db	1106	QDLVGHAPQCARSLTPTCT-CGSSDLYLVTR--HADVI	PVRRRGDSGLSPRISYLK 1162
Qy	97	TTSGFVSILFAPGAKQNEHTVTC-GAAATTSGLTSL--	FSPGASQNIQLITS--TDNSSP 151
Db	1163	GSSGPLELCPSG-----HVWGI	FIFAACVTRGAKAVDFVPVSMETTTWRSVPVTDNSSP 1216
Qy	152	PVPSFQVAHLHAPTGGSKTKVPAVAAQCYKVLVNLNPSVAATL	GFAYMSKAHGIDP 211
Db	1232	PAVQAFQVAVLHAPTGGSKTKVPAVAAQCYKVLVNLNPSVAATL	GFAYMSKAHGIDP 1276

212 NIRTGVRITTTGSPITVSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDDA 271  
1277 NIRTGVRITTTGAPITVSTYKFLADGCGSGAYDIIIMDCHECHSTDSTTILGIGTVLDDA 1336  
272 ETAGARLVVLTATPPGCVTPPHNIEEVALSTTGEIPFYKAIPLKGVKGRHLIFCHS 331  
1337 ETAGARLVVLTATPPGCVTPPHNIEEVALSTTGEIPFYKAIPIETIKGRHLIFCHS 1396  
332 KKKDELAALKLVGINAVAYRGLDVSVIPTSGDVVVVATDALMTGVTGDFSDVIDCNT 391  
1397 KKKDELAALKLVGINAVAYRGLDVSVIPASGNVVVATDALMTGVTGDFSDVIDCNT 1456  
392 C----- 392  
1457 CVTQTVDPSLDPTFTIETTTMPQDAVSRORRGRTRGRGIYRVFTPGRPSGMFDDSV 1516  
393 ----- 392  
1517 LCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVQCDHLEFWESVFTGLTHIDAHFLSQ 1576  
393 ----- 392  
1577 KOAGDNFFYLVAQTVCARAQAPPPSDQWQKCLTRLKPTLHGPTPLLYRLGAVQNEVT 1636  
393 -----AC-----SGKPAIIP 402  
1637 LTHPTIKYIMACSADLEVTSTWLVGVLAALAAAYCLTTGVSIVVIGRIILSGKPAVVP 1696  
403 DREVLYRFDEMEECQHLPIYEQGMALAEQFKQKALGLSGRGGPAIVDPKEVLYQQYDE 462  
1697 DREVLYQFDEMEECASHLPYIEQGMALAEQFKQKALGLQ-----TA 1739  
463 MEECSQAAPYIEQAQ-----VIAHQFKEKVLGILNDQVV----- 497  
1740 TKQAEAAAPVSVESKVRTLEAFWANDMWNFIQIQLAGLSTLPGPAPASLMAFTASITS 1799  
498 -VTPDKELY-----EAFDEMEECASKAALIERGORMAEMLSKIQGL 539  
1800 PLTTQSTLLFNILGGVAAQLAPPGAASAFVGCAGIAGAAGSIGLGVLDVMDVAGYGAGV 1859  
540 LG-----ILRRHVGPGEVGMNRL 560  
1860 AGALVAFKVMGSEMPSTEDLVNLLPAILSPGALVGVVCAAILRRHVDPGEGAVQMNRL 1919  
561 IAFASRGNHVSPTHVPS----- 578  
1920 IAFASRGNHVSPTHVPSDAAARVTQILSLTITOLLRLHQNEDCSTPCSGSLRD 1979  
579 ----- 578  
1980 VWDWICTVLADFKTWLQSKLLPRLPGVPFFSCQGYKGVWRGDGIMQITPCGAQLTGHV 2039  
579 ----- 578  
2040 KNGSMRIWGPKTCNTWHTGTPPINAYTTGCTPSPAPNYSRALWRVAEEYVEVRRVGD 2099  
579 -----RSRRFAQA----- 586  
2100 HYVTGMTDNVRCPCQVPAPEFTEVDGVRLHRYAPACKPLLRREEVSFQGLNQYVGSQ 2159  
587 LPVWARPD-----Y 595  
2160 LPCEPEPDVAVLTSMLTDPDASHITAEAKRRLARGSPPSLASSASQLSALKAACTTRH 2219  
596 NPP---LVET---WKK-----PDYEP-----PVHGRSSR 619  
2220 TTPDADLIEANLLRQEMGNITRVSEKNVILDSFDPLRAEEDEREVSVPAILRKSR 2279  
620 RFAQALPVWARPDYNNPLVETWKKPDYPPVHG 653  
2280 KPFPALPVWARPDYNNPLLEPNKDPDYPPVHG 2313

RESULT 8

GNWVCU

Genome polyprotein - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C:Accession: A39253; PS0086  
R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A:Reference number: A39253; MUID:91088550; PMID:2175903  
A:Accession: A39253  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <KAT>  
A:Cross-references: UNIPROT:P26662; GB:D90208; NID:G221610; PID:BAA14233.1; PID:G221611  
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 55B, 219-223, 1989  
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari  
A:Reference number: PS0085  
A:Accession: PS0086  
A:Molecule type: genomic RNA  
A:Residues: 2650-2707 <KA2>  
A:Experimental source: Japanese isolate  
A:Comment: The cleavage sites of this polyprotein have not been determined.  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <BPM>  
F:192-389/Product: major envelope protein E #status predicted <MSE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2

Query Match 34.1%; Score 1519; DB 1; Length 3010;

Beat Local Similarity 32.0%; Pred. No. 7.2e-86;

Matches 386; Conservative 44; Mismatches 108; Indels 670; Gaps 18;

QY 115 HVTG-GAARTSGTSL--PSPGASQNIQLTS--TDNSPPVVPVQSFQVAHLHAPTGS 169

DB 1175 HVGIFRAAVCTRGVAKAVDFIPVESMETTWRSVPVTDNSPPVPAVPTQFQVAHLHAPTGS 1234

QY 170 GKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDFNIRTGVRTITGSPITYS 229

DB 1235 GKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDFNIRTGVRTITGSPITYS 1294

QY 230 TYGFLADGGCGGAYDIIICDECHSTDATSIILGIGTVLDDAETAGARLVLTATPPGS 289

DB 1295 TYCKFLADGGCGGAYDIIICDECHSTDTSTILGIGTVLDDAETAGARLVLTATPPGS 1354

QY 290 VTVPHNPTEEVALSTTGEIPFYKAIPLKGVKGRHLIFCHSKKCKDELAALKLVGINA 349

DB 1355 ITVPHNPTEEVALSTTGEIPFYKAIPIEATKGGHRLIFCHSKKCKDELAALKLVGLNA 1414

QY 350 VAYYRGLDVSVIPTSGDVVVVATDALMTGVTGDFSDVIDCNTC----- 392

DB 1415 VAYYRGLDVSVIPTSGDVVVVATDALMTGVTGDFSDVIDCNTCTVQTVDPSLDPTFTI 1474

QY 393 ----- 392

DB 1475 TTLPODAVSRAQRGRGRTGRSGIYRVFTPGRPSGMFDSVLCBYCAGCAWYELTPAE 1534

QY 393 ----- 392

DB 1535 TSVRLRAYLNTPLPVQCDHLEFWESVFTGLTHIDAHFLSQTKQAGDNLPLYLVAYQATVC 1594

Qy	393	-----AC-----	394
Db	1595	ARAQAPPPSMDWMKCLRLKPTLHGPTPLLYRLGAVQNEVTLTHPTIKYIMACMSADLE	1654
Qy	395	-----SGKPAIPDPREVLYRFDMEECSEQH	420
Db	1655	VVTSTWLVGVLAALAAAYCLTTGTSVVIVGRILLISGRPAVIPDPREVLYQDFDEMEECASH	1714
Qy	421	LPYIEQGMMLAEQPKQKALGLSRGKPAIPVDPKEVLYQDYDEMEECSSQAAPYIEQAQ---	477
Db	1715	LPYIEQGMMLAEQPKQKALGLLQ-----TATKQAAAPAVVESKWRAL	1757
Qy	478	-----VIAHQPKFKEVLGLINDQV-----VTPDKELLY-----	506
Db	1758	EVFWAKHMMNFISGIQYLAGLSTLPGNPAIASLMAFTASITSLTTQNTLLFNILGGMVA	1817
Qy	507	-----EAPDEMEECASKAALIEGORMAEMLSKIKOGLLG-----	541
Db	1818	AQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYGAGVAGALVAFKVMGSEMPSTE	1877
Qy	542	-----ILRRHVGPGEGAVOMMRLLIAFASRGNHVSPTHYVP-	577
Db	1878	DLVNLPAILSPGALVGVVCAAILRRHVGPGEGAVOMMRLLIAFASRGNHVSPTHYVPE	1937
Qy	578	-----SRRRFAQALPV-----WAPDPYNPLPVETWKK-----	605
Db	1938	SDAAARVTQILSSITITQLLKRHLQWINEDCSTPCSGSLWKVDWDMICTVLSDFKTWLQS	1997
Qy	606	-----PDY-----	605
Db	1998	KLLPRLPGLPFLSCQGYKGVWRGDGIMQTTCPGCAQITGHVKNGSMRIVGPKTCSNTWH	2057
Qy	606	-----PDY-----	608
Db	2058	GTPPINAYTTCPTSPAPNYSRALWRVAEEYVEVTRVGDHFHYVTGTTDNVKKPCQVP	2117
Qy	609	-----EPPVV-----	613
Db	2118	APFEFTEVDGVRHLRYAPVCKPLLRBEVWFQVGLNQVLVGSQPCPEPDPVAVLTSMLTD	2177
Qy	614	-----HG-----	615
Db	2178	PSHITAETAKRLARGSPPSLASSASQLSAPSLKATCTTHHDSFPDADLIEANLLWRQEM	2237
Qy	616	-----RSSRFAQALPVWARPDYNPPL	637
Db	2238	GGNITRVESENKVVILDSFDPPIRAVEDEREISVPAETILRPRKPPPALPIWARPDYNPPL	2297
Qy	638	VETWKPKPDYPPVHVGRKTKRNTNRRPQDKFPGGQIVGGVYLLPRRGRPLGLVLRKT	697
Db	2298	LESWKDPDYPPVHVHG-----CP-----LPSTKA	2321
Qy	698	SPIPKARR	705
Db	2322	PIPPPPRR	2329

RESULT 9

QJ1303  
genome polyprotein - hepatitis C virus (isolate HC-J6)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: QJ1303  
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Miyakawa, Y.  
J. Gen. Virol. 72, 2697-2704, 1991  
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
A;Reference number: QJ1303; MUID:92044440; PMID:1658196  
A;Accession: QJ1303  
A;Molecule type: genomic RNA  
A;Residues: 1-3033 <OKA>

A;Cross-references: UNIPROT:P26660; GB:D00944; NID:g221650; PID:BAA00792.1; PID:g221651  
A;Experimental source: isolate HC-J6 from a Japanese individual  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmembrane  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1011-1619/Product: hepatitis C virus nonstructural protein NS2 #status predicted <NS2>  
F;1316-1321/Region: nucleotide-binding motif B  
F;1320-1323/Region: DEXH motif  
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 234, 305, 325, 417, 423, 430, 448, 477, 534, 542, 558, 578, 627, 649, 1091, 1217, 1259, 2038, 281

Query Match 32.1%; Score 1429.5; DB 1; Length 3033;  
Best Local Similarity 31.2%; Pred. No. 2.7e-80;  
Matches 393; Conservative 62; Mismatches 165; Indels 639; Gaps 27;

Qy	26	SNGPV-KVMGSIKGLTEGLHGFHFHFEFGDNTAGCTSAGPHFNPLSTRGCNCSTYVGHITG	84
Db	1097	SRGPVTQMYSASG---DLVGFSPGCTKSELPCTCGVDLY-LVTR--NADVIPARRRG	1150
Qy	85	HRMAWKIG-----SAARTTSGFVSLFAPGAKQNEHTVGTGAAARTTSGTSLFSFGSQNI	140
Db	1151	DKRGALLSPRLSTLKGSGGVLCPRG-----HAGVVFRA-----AVCSRGVAKSI	1197
Qy	141	QLI-----TSTNNSPPVVPQSFQVLAHLHAPTCGSGKSTKVPAYAAQGYKVLV	188
Db	1198	DFIPVETLDIVTRSPTFSDNSTPPAVPQTVQVGLHAPTCGSGKSTKVPAYAAQGYKVLV	1257
Qy	189	LNPSVAATLGFAYMSKAHGDINRTGVTITTTGSPITTYSTYKGLADGCGSGAYDII	248
Db	1258	LNPSVAATLGFAYMSKAHGINRTGVTITTTGSPITTYSTYKGLADGCGSGAYDII	1317
Qy	249	ICDECHSTATSLIGTGTVDLQDAETAGARLVLAATATPPGSVTVPHNIEVALSTTGEI	308
Db	1318	ICDECHAVDSTTILIGTGTVDLQDAETAGVRLTVLATATPPGSVTTPHNIEVALQGEI	1377
Qy	309	PFYGAIPLEVIKGGHLLFCHSKKCDLAALVALGINAVAYYRGLDVSIVPTSDVV	368
Db	1378	PFYGRAIPLSYIKGGHLLFCHSKKCDLAALRGMLNNAVAYYRGLDVSIVPTQGDVV	1437
Qy	369	VVATDALMTGTCDFDSVDCNT-----	391
Db	1438	VVATDALMTGTCDFDSVDCNVAVTVQVDFSLDPTFTTTTQVPODAVSRQRGTGR	1497
Qy	392	-----CAC--SG-----KPAIIPDREVLYREFDEMEECSEQ	419
Db	1498	GRLGIVYVSTGERASGMFDSVVLCEYDAGAAWYELTPAETTVRLRAYENTPGLPVCQD	1557
Qy	420	HLP-----YIEQGMMLAEQF-----KOKALGLSRGG-----	445
Db	1558	HLFEFWEAVFTGLTHIDAHFLSQTQKSGENFAYLTAYQATVCARAKAPPPSMDVMMKCLTR	1617
Qy	446	-KPAIVDPKEVLYQDYDEMEECSSQAAP---YISQ-----AOVIAH	481
Db	1618	LKPTLVGPTPLLYRLGSGVTVNEVTLTHPTIKYIMACMSADLEVMTSTWVLAGGVLAQVAA	1677
Qy	482	QFKEKVLGLJD-----NDQVVVTPDKELIYFAFDEMEECASKAALIEEGORMAEMLSKIQ	537
Db	1678	CLATGCVCIIGRLHVNQRAVAVPDKVLYEAFDEMEECASRAALIEEGORIAEMLSKIQ	1737
Qy	538	GLL-----	540
Db	1738	GLLQASQKQADIQPAVQASWPKEQFWAKHMMNFISGIQYLAGLSTLPGNPAVASMMAF	1797
Qy	541	-----	540
Db	1798	SAALTSPSTSTTILLNILGGLASQIAPPAGATGTFVVSGLVGAANVGSIGLGVLDILA	1857

```
Qy 541 -----GILRRHVGPGEAV 554
Db 1858 GYGAGISGALVAFKIMSGEKSPMEDVVNLLPGILSPGALVVGVCIAILRRHVGPGEAV 1917
Qy 555 QMWNRLIAFASRGNHVSPTHV--PKSRFAQAL-----587
Db 1918 QMWNRLIAFASRGNHVAPTHVTTESDASQRTQLGLSLTITSLRLHNWITEDCIPCS 1977
Qy 588 -----PVNA-----591
Db 1978 GSWLRDWDVVCITLTDKWLTKLPFKMPLPFISCGKYGKVGWAGTGIMTTRCPCGA 2037
Qy 592 -----RPNPLVETWK-----604
Db 2038 NISGNVRLGSMRITGPKTCMNIMOGTTPINCYTEGQCVPKPAFNK---IAIWRVAASEY 2094
Qy 605 -----KP-----606
Db 2095 AEVTHQSGSYHITGLTDNLKVPCLPSPFFSWDGVQIHRFAPIPKPFPRDEVSPCVG 2154
Qy 607 -----DYPE-----610
Db 2155 LNSFVVGSQLPCDEPDTDLVLTSMIDPSHITAETAAARLARGSPPEASASSASQLSAPS 2214
Qy 611 -----PVHGRS-----617
Db 2215 LRATCTTHGKAYDVMVDANLFMGDVTRIESESKVVLVDSLDPMEVERSLSPSPSEY 2274
Qy 618 ---SRRAQLPWARDVNPPLVETWKDYPVPHG-----RKTENTRRPODV 667
Db 2275 MLPKGRFPALPAWAPDYNPPLVESWKRPDYQAPVAGCALPPEKPTPTPPRRRRRTV 2333

RESULT 10
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contents: capsid protein C; envelope protein M; hepatitis C virus (nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A40250; PQ0397; F00559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repc
A;Reference number: A40250; UID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:G221608; PIDN:BAA01761.1;
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I.
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; UID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; UID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BAA01418.1; PID:G221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
P;1-115/Product: capsid protein C #status predicted <CPC>
P;116-191/Product: envelope protein M #status predicted <EPM>
P;192-389/Product: major envelope protein E #status predicted <MEE>
P;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
P;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
```

```
F;1011-1619/Product: hepatitis C virus #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEX motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23;
```

Query Match 31.9%; Score 1422.5; DB 1; Length 3033;

Best Local Similarity 29.6%; Pred. No. 7.4e-80;

Matches 408; Conservative 62; Mismatches 182; Indels 725; Gaps 31;

Qy 22 EQKESNGPVKWSIKGLTEGLHGFV-----HEFGDNTAGCTAGPHPLSTRGC 73

Db 1055 DKNEAQGVQLSSV---TQTLGTSISGLVMTVYHAGNKTLL---AGP-----KGP 1100

Qy 74 NCSIIYPGHITGHRMANKIGSAART---TSGFYSLF-----AP 107

Db 1101 VTQMYTS-AGDVLVGPSPPTKSLDPCTCGADVLLVTRNADVIPRRKDDRRGALLSP 1159

Qy 108 GAKQNETHTVTGAAARTTSLTSLF-----SPGASONIQLI-----TSTDNSS 150

Db 1160 RPLSTLKGSSGGPVLCRSHGAVGLFRAAVCAVCAKSIDFIPVESLDVATRTSPFSNST 1219

Qy 151 PPVVPQSFQVAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGID 210

Db 1220 PPAPVQSYQVGYLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGIN 1279

Qy 211 PNRTGTVRTTITTSPTTYSTYTGKFLADGGCGGAYDIIICDECHSTDATSLTGLGTVDQ 270

Db 1280 PNRTGTVRTTITTSPTTYSTYTGKFLADGGCGGAYDIIICDECHSDVATTLTGLGTVDQ 1339

Qy 271 AETAGARLVVLATATPPGVSIVTPHPNIEEVALSTGTGIPFYKAIPLKVIKGRHLIFCH 330

Db 1340 AETAGARLVVLATATPPGVSIVTPHPNIEEVALSTGTGIPFYKAIPLKVIKGRHLIFCH 1399

Qy 331 SKKKDBELAAKLVALGNVAVAYRGLDVSVITPSGDVVVATDALMTGYGDFSVSDCN 390

Db 1400 SKKKDBELAAKLVALGNVAVAYRGLDVSVITPSGDVVVATDALMTGYGDFSVSDCN 1459

Qy 391 TC-----ACSG-KPAIIPDRE 405

Db 1460 VAVSQIVDFSLDPTTITTTQTVQDVAVSQRGRTGRLGVYRVYSSSERPSGMDSV 1519

Qy 406 VLYREFDE-----MEECQHLPIYIE---QGMW-----LAEQ 433

Db 1520 VLCEYDAGAAWYELTPAETTVRLRAYFNTPLGVPVCDHLEFWEAVFTGLTHIDAHFLSQ 1579

Qy 434 FKQKALGLSRGG-----KPAIVDPKEVLYQQ 459

Db 1580 TKQ-----GGENFAVLTAYQATVCAKAKAPPSPSDVMWKKLTRLKPTLTGTPTLLYRL 1632

Qy 460 YDEMECSQAAP---YIEQ-----AQVIAHQFKEKVLGLID---492

Db 1633 GAVTNEVTLTHPTVKYIATCQADLEIMTSSWSVLAVGVAAYATCATGCISIIIRLHL 1692

Qy 493 NDQVWVTPDKIELYEADEFDEMECSKAALIEEGORMAEMLSKTIQGLL-----540

Db 1693 NDRVVAAPDKIELYEADEFDEMECSKAALIEEGORMAEMLSKTIQGLL-----1752

Qy 541 -----540

Db 1753 AIQSSWPKLEQFWAKHWNFTISGTYIYGLAGLSTLPGNPAVAMMAFSAALTSPLTSTIL 1812

Qy 541 -----540

Db 1813 LNIIMGWLASQIAPPAGATGFVWSGLVAAVSGTGLKILVDVLVAGVAGISGALVAPKI 1872

Qy 541 -----GILRRHVGPGEAVQVNMNRLIAFASRGNH 569

Db 1873 MSGKPTVEDVNNLLPAILSPGALVVGVCIAILRRHVGPGEAVQVNMNRLIAFASRGNH 1932





Db 2199 LANSASQSLAPSLKATCTIOGHHDDADLIKANLLWRCQMGNTTRVEAKNKVILDCFK 2258  
Qy 621 -----FAQALPVWAPDYNNPLVETWKKPDYEPVHVGRKTK 657  
Db 2259 PLKEEDDDRETSVADCFKGPAPFPALPVWAPGYDPPLETKWRPDYDPPQWGPC-- 2316  
Qy 658 RNTNRRPDQVFPGGGQIVGVYLLPRGPRGLGVLATRKTSPI 700  
Db 2317 -----IPPAGPPPVPLPRKRKXPM 2335

RESULT 12  
S68016  
ATPase/RNA helicase - hepatitis C virus (fragment)  
C:Species: hepatitis C virus  
A:Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: S68016  
R:Jin, L.; Peterson, D.L.  
Arch. Biochem. Biophys. 323, 47-53, 1995  
A:Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
A:Reference number: S68016; MUID:96019946; PMID:7487072  
A:Accession: S68016  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-386 <JIN>  
A:Cross-references: UNIPROT:Q04045  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polyprotein  
F:24-31/Region: nucleotide-binding motif A (P-loop)  
F:86-91/Region: nucleotide-binding motif B  
F:90-93/Region: DEXH motif

Query Match 25.3%; Score 1128.5; DB 2; Length 386;  
Best Local Similarity 89.1%; Pred. No. 7.9e-63;  
Matches 221; Conservative 4; Mismatches 2; Indels 21; Gaps 2;  
Qy 146 TDN-SSPPVQSPQFVAHLHAPTSGKSTKVPAAAGYKVLVNLNPSVAATLGFAYMS 204  
Db 4 TDNSSPPVQSPQFVAHLHAPTSGKSTKVPAAAGYKVLVNLNPSVAATLGFAYMS 63  
Qy 205 KAHGIDPNIRTCVRIITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGI 264  
Db 64 KAHGYD-----YKFLADGCGSGAYDIIICDCHSTDATSIIGI 103  
Qy 265 GTVLDAQETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVIEKGR 324  
Db 104 GTVLDAQETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVIEKGR 163  
Qy 325 HLIFCHSKKCDLAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGTGDFD 384  
Db 164 HLIFCHSKKCDLAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGTGDFD 223  
Qy 385 SVIDNCTC 392  
Db 224 SVIDNCTC 231

RESULT 13  
PC2219  
polypeptide - hepatitis C virus (type 5a) (fragments)  
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote  
C:Species: hepatitis C virus  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: PC2219  
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the H  
A:Reference number: PC2219; MUID:94338342; PMID:7520237  
A:Accession: PC2219  
A:Molecule type: mRNA  
A:Residues: 1-876 <STU>  
A:Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579  
A:Experimental source: serum

C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Glycoprotein  
F:1-191/Product: core #status predicted <COE>  
F:68-78/Region: variable  
F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F:248-338/Region: E2  
F:339-411/Region: NS1 (amino end)  
F:412-783/Product: NS3 #status predicted <NSR>  
F:784-837/Product: NS4 #status predicted <NSA>  
F:838-876/Product: NS4B #status predicted <NSB>  
F:281,287,294,312,340/Binding site: carbohydrtate (Aan) (covalent) #status predicted

Query Match 21.4%; Score 954.5; DB 2; Length 876;  
Best Local Similarity 34.6%; Pred. No. 1.7e-51;  
Matches 252; Conservative 57; Mismatches 146; Indels 273; Gaps 18;  
Qy 32 VWGSIKGLTEGL-HGFHVHEFGDNTAGCTSGAGPHFNPLSTRGCNCIYPGH----- 81  
Db 140 VGGPIGVARALANGVRVLEGVNAYTGNLPGCSFSFILALLSCLTVPASAPTTALLVA 199  
Qy 82 -----ITGHRMAWK-----GSAATTSGFVSLFAPGAKQKETHVTGGAAA 122  
Db 200 QLLRIPQVVIDIIAGSH--WGVFAAAYASVANWTKVVLVFLFAGVDATTQISGSSA 257  
Qy 123 RTTSGTSLSPGASQNIQLITSD-----NSSPPVQSPQVA-----HLHAPTGSCKS 172  
Db 258 QTTTGIASTFTRGAQQLQTLINTNGSMHINRTALNCNDSLOTGTGFIAGLYFYHKFNSSGCP 317  
Qy 173 TKVPA---AYAAQGYKVL---VLNPS-----VAA 195  
Db 318 DRMASCRALAFDDQWGTISYANISGFSDDKPYCHVPPRCGVVPAQVCGPVYCFTPS 377  
Qy 196 TLGFAGYMSKAHGDIPNIRTCVRIITGSPITYSTYCKFLADGCGSGG 243  
Db 378 PVVVGTTDSKGH---PTVNWGSNVTDFPLMNTNRPPIITGASITYSTYCKFLADGCGSGG 434  
Qy 244 AYDIIICDCHSTDATSIIGITGVLDOAETAGARLVVLTATPPGVTVPHPNIEVALS 303  
Db 435 AYDIIICDCHSQDATTILGIGTGLDQAETAGARLVVLTATPPGVTVPHPNIEVALP 494  
Qy 304 TTGEIPFYKAIPLVIEKGRHLIFCHSKKCDLAKLVALGINAVAYRGLDVSVIPT 363  
Db 495 QEGEVFPFYGRAIPLAFIKGGRHLIFCHSKKCDLAKOLTSLGNNAVAYRGLDVAVIPT 554  
Qy 364 SGDVVVVATDALTMTGTGDFSDVIDCNTC----- 392  
Db 555 AGDVVVCSTDALMTGTGDFSDVIDCNSAVTQTVDFSLDPTFTTETTTVPQDAVSRQR 614  
Qy 393 -----ACSGKPAIIPDREVLYREFDE-----M 414  
Db 615 GRTGRGRHGIYRVVSAGERPSDMFDSVVLCECYDAGCAWYDLTPAETTVRLRAYINTPGL 674  
Qy 415 EECQHLPLYE-----QG-----WMLAQ-----EKOKA----- 438  
Db 675 PVCQDHLFEWEGVFTGLTNIDAHMLSQTKQGENPPYLVAYQATVCVRAKAPPSPSWDTMW 734  
Qy 439 -----LG----- 440  
Db 735 KCMLEKLTLTGPTPLLYRLGPVQNEITLTHPIITKYIMACMSADLEVITSTWLVGVVVA 794  
Qy 441 -----LSRGGKPAIVDPKKEVLVYQYDEMEECSQAAPYIEQAQVIAHQF 483  
Db 795 ALAAVCLTVGSVAIVGRIILSGKPAIIPDREALYQOFDEMEBECASLPYMDETRAIGOF 854  
Qy 484 KEKVLGLI 491  
Db 855 KEKVLGFI 862

RESULT 14  
S21337  
genome polyprotein S4 (NS3 region) - hepatitis C virus (fragment)



C;Species: hepatitis C virus  
C;Date: 20-Feb-1995 #sequence\_revision 19-Apr-1996 #text\_change 09-Jul-2004  
C;Accession: S21337  
R;Sato, A.  
submitted to the EMBL Data Library, April 1992  
A;Description: A sensitive serodiagnosis of hepatitis C virus infection with two cloned  
A;Reference number: S21336  
A;Accession: S21337  
A;Molecule type: Genomic RNA  
A;Residues: 1-216 <SAT>  
A;Cross-references: UNIPROT:Q68968, EMBL:X65547, NID:g59494, PID:CAA46516.1; PID:g59495  
C;Superfamily: hepatitis C virus genome polyprotein  
C;keywords: polyprotein

Query Match 19.3%; Score 860; DB 2; Length 216;  
Best Local Similarity 94.7%; Pred. No. 1.8e-46;  
Matches 162; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
Qy 222 TGSPIYTVYKGLADGGCGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVWL 281  
Db 1 TGAPITYTVYKGLADGGCGGAYDIIIMCDECHSTSTSLGIGTVLDQAETAGARLVWL 60  
Qy 282 ATATPPGSVTVPHNIEEVALSTGEIPFYKKAIPLEVIKGRHLIFCHSKKKDELA 341  
Db 61 ATATPPGSVTVPHNIEEVALSNTGEIPFYKKAIPETIKGRHLIFCHSKKKDELA 120  
Qy 342 LVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTYGDFSDVIDCNTC 392  
Db 121 LSALGLNAVAYRGLDVSVIPTSGDVVVVATDALMTGTYGDFSDVIDCNTC 171

## RESULT 15

T08841  
polyprotein - douroucouli hepatitis GB virus A  
C;Species: douroucouli hepatitis GB virus A  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C;Accession: T08841  
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
J. Gen. Virol. 79, 41-45, 1998  
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
A;Reference number: Z16486; MUID:98120818; PMID:9460920  
A;Accession: T08841  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3005 <ERK>  
A;Cross-references: EMBL:AF023425; NID:g2828599; PID:AAC40502.1; PID:g2828600  
C;Superfamily: hepatitis C virus genome polyprotein  
C;keywords: polyprotein

Query Match 13.6%; Score 608; DB 2; Length 3005;  
Best Local Similarity 37.0%; Pred. No. 3.6e-29;  
Matches 155; Conservative 54; Mismatches 142; Indels 68; Gaps 14;  
Qy 22 EQKESNGPVKWSIKGLTGG--LHG-----FHVHEFGDNTAGCTSAGPHNPLSTR--- 71  
Db 1002 DEREHGSIVVLGTSTTRSGTCVNGVYMTTFH-----GSNAR--TLAGP-VGPVNCRWWS 1054  
Qy 72 -----GCNCSIYPGHI TCHRMWKLGSAA RTSG-----FVSLFAP-- 107  
Db 1055 PSDDVAVYPLPSGASCLPECKC-----GTQSVWCIRNDGALCHGRLSKLVLDLPTE 1106  
Qy 108 -----GAKQ-----NETHVTGGAARTTSG--LTSLSFGASQNIQLITSTDNSSPPVWP 155  
Db 1107 ISDFRGSSGSPILCDEGHVVGMMVSVLHRGVKVTGVYVVKPWETLPKDSQVKSEAPPVPG 1166  
Qy 156 QS-FQVAHLHAPTCGSGKTKVPAAYAAQGYKVLVNPVAAATLGFAYMSKAGIDPNIR 214  
Db 1167 KTGTEAPLYLPTGSGKSTRIPLEYTKAGHKVVLNPSIATVRAMGPFMEKLSQHPSIY 1226  
Qy 215 TGVRTI----TTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDQ 270  
Db 1227 CGHDTTAYSTTGSPLYCTYGRFMANPRRYLRGADIVI CDECHVTDPTSVLGMGRARLL 1286

Qy 271 AETAGARLVVLATATATPPGSVTVPHNIEEVALSTTGEIPFYKKAIPLEVIKGRHLIFCH 330  
Db 1287 ARECGVRLLLFATATPPGAPLAQHESI KEVPLGVGDGEVAFYGHKLPVERYRTGRHLLFCH 1346  
Qy 331 SKKKCDELA AKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGTYGDFSDVIDC 389  
Db 1347 SKVECNRLHAALSTAGCNNAVYYIRGNEQEI--PAGDVCVCATDALSTGYTGGFSTVTD 1403

Search completed: November 7, 2005, 20:11:17  
Job time : 41.1255 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 106.241 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-4  
Perfect score: 4455  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRTGSKGKGPVWP 829

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1624.5	36.5	2436	2 Q81756	Q81756 hepatitis c
2	1624.5	36.5	3011	1 POLG HCV1	P26664 h genome po
3	1620.5	36.4	3011	2 Q91FE5	Q91FE5 hepatitis c
4	1597.5	35.9	3011	2 Q36579	Q36579 hepatitis c
5	1597.5	35.9	3011	2 Q36608	Q36608 hepatitis c
6	1597.5	35.9	3011	2 Q36610	Q36610 hepatitis c
7	1597.5	35.9	3015	2 Q9PMU9	Q9PMU9 hepatitis c
8	1597.5	35.9	3015	2 Q9PMU5	Q9PMU5 hepatitis c
9	1596	35.8	3011	2 Q03463	Q03463 hepatitis c
10	1590.5	35.7	3011	2 Q36609	Q36609 hepatitis c
11	1586.5	35.6	2908	2 Q61X04	Q61X04 hepatitis c
12	1583.5	35.5	3011	2 Q9ELS8	Q9ELS8 hepatitis c
13	1574	35.3	3011	2 Q9DIT6	Q9DIT6 hepatitis c
14	1559.5	35.0	3011	1 POLG HCVH	P27958 h genome po
15	1559	35.0	3010	2 Q9DTE8	Q9DTE8 hepatitis c
16	1546	34.7	3010	2 Q68788	Q68788 hepatitis c
17	1546	34.7	3010	2 Q81757	Q81757 hepatitis c
18	1546	34.7	3013	2 Q6J6P5	Q6J6P5 hepatitis c
19	1545	34.7	3010	2 Q9Q1Y1	Q9Q1Y1 hepatitis c
20	1544.5	34.7	3010	2 Q9J3G4	Q9J3G4 hepatitis c
21	1544.5	34.7	3010	2 Q9J3H0	Q9J3H0 hepatitis c
22	1544	34.7	3010	2 Q9J3G5	Q9J3G5 hepatitis c
23	1542.5	34.6	3010	2 Q9DTE7	Q9DTE7 hepatitis c
24	1542.5	34.6	3010	2 Q9Q1X1	Q9Q1X1 hepatitis c
25	1542.5	34.6	3010	2 Q9Q1X2	Q9Q1X2 hepatitis c
26	1541.5	34.6	3014	2 Q9DTE0	Q9DTE0 hepatitis c
27	1541	34.6	3010	2 Q9Q1Y2	Q9Q1Y2 hepatitis c
28	1539.5	34.6	3010	2 Q9WMX2	Q9WMX2 hepatitis c
29	1539	34.5	1984	2 Q7T4V8	Q7T4V8 hepatitis c
30	1538.5	34.5	3010	2 Q9J3G6	Q9J3G6 hepatitis c
31	1538.5	34.5	3010	2 Q9J3H5	Q9J3H5 hepatitis c

RESULT 1

ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M32084; AAA45677.1; -			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P27958; 1A1V.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0019028; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008235; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006508; F:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006350; P:transcription; IEA.			
DR	GO; GO:0019079; P:viral genome replication; IEA.			
DR	GO; GO:0019087; P:viral transformation; IEA.			
DR	InterPro; IPR000345; CytC_heme_BS.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002531; HCV NS1.			
DR	InterPro; IPR000745; HCV NS4a.			
DR	InterPro; IPR001490; HCV NS4b.			
DR	InterPro; IPR002868; HCV NSa.			
DR	InterPro; IPR002166; HCV RdRp.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR009003; Peptidase_S29.			
DR	InterPro; IPR002518; Pept_U39_HCV NS2.			
DR	InterPro; IPR007095; RNA_pol_DS_PS.			
DR	InterPro; IPR007094; RNA_pol_PSVir.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	Pfam; PF01538; HCV NS2; 1.			
DR	Pfam; PF02907; HCV NS3; 1.			
DR	Pfam; PF01006; HCV NS4a; 1.			
DR	Pfam; PF01001; HCV NS4b; 1.			
DR	Pfam; PF01506; HCV NS5a; 1.			

32	1538.5	34.5	3010	2	Q9Q1Y3	Q9qiv3 hepatitis c
33	1536	34.5	3010	2	Q9DTE9	Q9dte9 hepatitis c
34	1535.5	34.5	3010	2	Q9Q1Y4	Q9qiY4 hepatitis c
35	1535	34.5	3010	2	Q9Q1Y9	Q9qiY9 hepatitis c
36	1535	34.5	3010	2	Q9Q1Z0	Q9qiZ0 hepatitis c
37	1534.5	34.4	3010	2	Q9DTE5	Q9dte5 hepatitis c
38	1534.5	34.4	3010	2	Q9J3G8	Q9j3g8 hepatitis c
39	1534.5	34.4	3010	2	Q9Q1X6	Q9qix6 hepatitis c
40	1534	34.4	3010	2	Q88949	Q88949 hepatitis c
41	1533	34.4	3010	2	Q9J3G2	Q9j3g2 hepatitis c
42	1533	34.4	3010	2	Q9Q1X7	Q9qix7 hepatitis c
43	1533	34.4	3010	2	Q9Q1X8	Q9qix8 hepatitis c
44	1532.5	34.4	3010	1	POLG HCVJT	Q00269 h genome po
45	1532.5	34.4	3010	2	Q807F3	Q807p3 hepatitis c

## ALIGNMENTS

DR Pfam; PF00271; Helicase C; 1.  
DR Pfam; PF00998; Viral RGR; 1.  
DR SMART; SMO0487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON TER 1  
FT NON TER 2436  
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 36.5%; Score 1624.5; DB 2; Length 2436;  
Best Local Similarity 36.1%; Pred. No. 1.7e-91;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;

QY 146 TDNSPPVQSFQVAVHLHAPTSGSKTKVPAAYAAQGYKVLVLPSPVAATLFGAYMSK 205  
DB 761 TDNSPPVQSFQVAVHLHAPTSGSKTKVPAAYAAQGYKVLVLPSPVAATLFGAYMSK 820

QY 206 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIG 265  
DB 821 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIG 880

QY 266 TVLDOAETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKKAIPLEVIKGRH 325  
DB 881 TVLDOAETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKKAIPLEVIKGRH 940

QY 326 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGVTGDFDS 385  
DB 941 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGVTGDFDS 1000

QY 386 VIDCNTC----- 392  
DB 1001 VIDCNTCVTQVDFSLDPTFTIETILPQDAVSRTRGRGKGYRFPVAPGERPSG 1060

QY 393 ----- 392

DB 1061 MFDSSVLCEYDAGCAWYELTPAETTVLRAYMNTPLGVPQDHLFEWEGVFTGLTHIDA 1120

QY 393 ----- 392

DB 1121 HFLSQTQSGENLPYLVAQYATVCARAQAPPPSWDQWKKLIRUKPTLHGTPPLLYRLGA 1180

QY 393 -----ACSG 396  
DB 1181 VQNEITLTPVTKYIMTMSADLEVTSTWLVGVGLAALAAAYCLSTGCVVIGRVLGS 1240

QY 397 KPAILPDREVLRYREFDEMECSQHLPIYIEQGMMLAEQFKOKALGL-----SRGKKAIPVD 452  
DB 1241 KPAILPDREVLRYREFDEMECSQHLPIYIEQGMMLAEQFKOKALGLLQTSRQAE-VIAPA 1299

QY 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEI 504  
DB 1300 VQTNWQKLETFWAKHMNFISGTYLQAGLSTLPG--NPATASLMFTAATVTSPTLTSQTL 1357

QY 505 LYE-----AFDEMECSKAALIEEGQMAEMLKSKIQLLGL----- 541  
DB 1358 LFNILGGVAAQLAAPGAATAFVGCAGLAGAIGSVGLGKVLIDILAGYGVAGVAGVAFK 1417

QY 542 -----ILRRHVGFEGAVQMMNRLLAFASRGN 568  
DB 1418 INSGEVPSTEDLVNLLPAILSPGALVGVVVCVCAAILRRHVGFEGAVQMMNRLLAFASRGN 1477

QY 569 HVSPTHVUPS----- 578  
DB 1478 HVSPTHVPSDAAARVTAISSLTVTQLRLRHQWISSECTPCSGSWLRDINDWICEV 1537

QY 579 ----- 578

DB 1538 LSDFKTLWAKALPQLPGIPFVSCQGYKGWVRVDGIMHTRCHGAEITGHVNGTWRIV 1597

QY 579 ----- 578

DB 1598 GPRTCRNMWSTGFPIINAYTTCPTPLPAPNTYTFALMRVSAEYVEIRQVGFHYVTGTTT 1657

QY 579 -----RSRRFA-----QALPVVWARPD 594  
DB 1658 DNLKCPQVPSPEFFTELDGVRLLHRRAPPCKPILLREVSFVGLHEYPVGSQLPCEPEPD 1717

QY 595 Y-----NPP-----LV 600  
DB 1718 VAVLTSMLTDPSSHITAEAGRRLLARGSPPSVASSASQSLKATCTATCTAHSDPDAELI 1777

QY 601 ET---WKK-----PDYEPVVHG-----RSRRFAQALPV 627  
DB 1778 EANLLARQMGNGNITRVESKNKVILDSFDPLVAEEDEREISVPAEILRKSRRAQALPV 1837

QY 628 WARPDYNPPLVETWKKPDPYEPVVHG 653  
DB 1838 WARPDYNPPLVETWKKPDPYEPVVHG 1863

RESULT 2  
POLG\_HCV1  
ID\_POLG\_HCV1 STANDARD; PRT; 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).  
OS Hepatitis C virus (isolate 1) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
RA Gallegos C., Cui D., Medina-Selby A., Barr P.J., Weiner A.J.,  
RA Bradley D.W., Kuo G., Houghton M.;  
RT "Genetic organization and diversity of the hepatitis C virus";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
CC -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA)(N).  
CC -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein C and mRNA.  
CC -I- SIMILARITY: Contains 1 peptidase S29 domain.  
CC -I- SIMILARITY: Contains 1 peptidase U39 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M62321; AAA45676.1; -.  
DR PUR; A39166; GNWVC3.  
DR PDB; 1HEI; X-ray; A/B-1206-1656.  
DR PDB; 1ONB; NMR; A-1349-1507.  
DR MEROPS; S29.001; -.  
DR MEROPS; U39.001; -.

[illegible]

QY 393 ----- 392  
Db 1511 MPDSSVLCYDAGCAYELTPAETTVRLRAYMNTPCLPVCQDHLPEWEGVFTGLTHIDA 1570  
QY 393 ----- 392  
Db 1571 HFLSOTKQSGENLPYLVAQYATVCARAQAPPSPDQWKKLIRLKLPTLHGTPLLYRLGA 1630  
QY 393 -----ACSG 396  
Db 1631 VQNEITLTHPVTKYIMTMSADLEVTWTWLVGGVLAALAAVCLSTGCVIVIGRVVLGS 1690  
QY 397 KPAAIIDREVLYREFDEMEECSQHLPIYIEQMMLAEOFKQKALGL-----SRGCKPAIVPD 452  
Db 1691 KPAAIIDREVLYREFDEMEECSQHLPIYIEQMMLAEOFKQKALGLLQATASRQAE-VIAPA 1749  
QY 453 KEVLYQOYD-----EMEECSQAAPYIEQAQVIAHQFKVKGLINDNDVVVTP-----DKEI 504  
Db 1750 VQTNWQKLETFWAKHWMNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSTQL 1807  
QY 505 LYE-----AFDEMEECASKAALIEEGORMAEMLSKIOGLLG----- 541  
Db 1808 LFNILGWAAQAAPGAATAFVAGLAGAAGISVGLKGLVLIDILAGYAGAGVAGALVAFK 1867  
QY 542 -----ILRRHVGPGEAGVQMMNRLIAFASRGN 568  
Db 1868 IMSGEVPTEDLVNLLPAILSPALVGVVCAAILRRHVGPGEAGVQMMNRLIAFASRGN 1927  
QY 569 HVSPTHYVPS----- 578  
Db 1928 HVSPTHYVPSDAAARVTAAILSLVTQLRLRHQWISSECTTPCSGSLURDIWDWICEV 1987  
QY 579 ----- 578  
Db 1988 LSDFKTLWAKLMPQLPGIPFVSCQYKGVWRVDMHTRCHGAEITHGVNKGTMRIIV 2047  
QY 579 ----- 578  
Db 2048 GPRTCRNMMSGTFPINAYTTGCTPLPAPNYTALMRVSAEYVEIRQVDFHVVGTMTT 2107  
QY 579 -----RSRFA-----QALPFWARP 594  
Db 2108 DNLKPCQVPSPEFFTELDGVLHREAPPCKLLREBVSFRVGLHVPVGSQPCPEPD 2167  
QY 595 Y-----NPP-----LV 600  
Db 2168 VAVLTSLTDPDSHITAEAGRLRAGSPSPVASSASQLSAPSLKATCTANHSDPAELI 2227  
QY 601 ET---WKK-----PDYPPVVHG-----RSSRFAQALPV 627  
Db 2228 EANLLWRQMGGNITRVESKENVILDSFDPLVAEEDEREISVPAELRLKSRFAQALPV 2287  
QY 628 WARPDPNPPLVETWKKPDYPPVVHG 653  
Db 2288 WARPDPNPPLVETWKKPDYPPVVHG 2313

RESULT 3

Q9IFES ID Q9IFES PRELIMINARY; PRT; 3011 AA.  
AC Q9IFES;  
DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21262212; PubMed=11369872;  
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;

Query Match 36.4%; Score 1620.5; DB 2; Length 3011;  
Best Local Similarity 36.0%; Pred. No. 4.1e-91;  
Matches 398; Conservative 27; Mismatches 80; Indels 601; Gaps 15;  
QY 146 TONSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205  
Db 1211 TONSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270  
QY 206 AHGIDPNIRTVRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATSILGIG 265  
Db 1271 AHGIDPNIRTVRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATSILGIG 1330  
QY 266 TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRH 325  
Db 1331 TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRH 1390

RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype  
sequence.";  
RT J. Gen. Virol. 82:1291-1297(2001).  
DR EMBL; AF271632; AAF81759.1; -.  
DR PIR; A44150; A44150.  
DR PIR; Q0804; Q0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; Q8JYS1; ICWX.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002186; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept Ser Cys.  
DR InterPro; IPR002518; Pept U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02307; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN.1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;

```
QY 326 LIFCHSKKCKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 385
Db 1391 LIFCHSKKCKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 1450
QY 386 VIDCNC-----392
Db 1451 VIDCNCVTOTVDFSLDPTTITITLPQAVSRTQRRGTGRGKPGIYRFVAPGERPSG 1510
QY 393 -----392
Db 1511 MFDSVLCECYDAGCANWELTPAETTVRLRAYMYNTPGLPVCQDHLFEWEGVFTGLTHIDA 1570
QY 393 -----392
Db 1571 HFLSOTKQSGENLPYLVAQATVCARAQAPPSWDQMKLIRLKPTLHGPTPLLYRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEITLHPVKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVVLSG 1690
QY 397 KPAIIPREVLVREFDEMECSQHLPIEQGMMLAEQFKQKALGL-----SRGGRPAIIVPD 452
Db 1691 KPAIIPREVLVQEFDEMECSQHLPIEQGMMLAEQFKQKALGLLQTSRQAE-VIAPA 1749
QY 453 KEVLYQDYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVVTP---DKEI 504
Db 1750 VQTNWQKLETFWAKHWNFISGIQLAGSLTPG--NPAIASLMAFTAATAVTSPLTTSQTL 1807
QY 505 LYE-----AFDEMECSKAALIEGQMAEMLKSKIQGLLG-----541
Db 1808 LFNILGGVAAQLAAPGAATAFVGAGLAGAIGSVGLKVLIDLAGVGAGVAGALVAFK 1867
QY 542 -----ILRRHVGPGEAGVQWNRRLIAFASRGN 568
Db 1868 IMSEVPSTEDLVNLLPAIILSPGALVGVVCAAILRRHVGPGEAGVQWNRRLIAFASRGN 1927
QY 569 HVSPTHVPS-----578
Db 1928 HVSPTHVPSDAAARVTAIILSLTVTQLRLRHQWISSECTTPCSGSLRDIWDWICEV 1987
QY 579 -----578
Db 1988 LSDFKTLWAKMLPQLPGIPFVSCQGYGVWRGDGIMHTRCHGAEITGHVKNGTWRIV 2047
QY 579 -----578
Db 2048 GPRTCRNWSGTFPINAYTTGCTPLPAPNTYFALWRVSAEYVEIRQVGFHYVTGTT 2107
QY 579 -----RSRFA-----QALPWARPD 594
Db 2108 DNLKPCQVPSPFEFFTELDGVLRLHRFAPCKPLLRREEVSRVGLHEYPVGSQLPCEPDP 2167
QY 595 Y-----NPP-----LV 600
Db 2168 VAVLTSLMTPSHITAEAGRLARGSPSVASSASQLSAPLTKATCTANHDSFDAL 2227
QY 601 ET---WKK-----PDYEPVPHG-----SSRRRFAQALPV 627
Db 2228 EANILMROEMCGNITRVESENKVVILDSFDPLVAEEDEREISVPAEILKRRRFAQALPV 2287
QY 628 WARDYNPPLVETWKKPDYEPVPHG 653
Db 2288 WARDYNPPLVETWKKPDYEPVPHG 2313
```

## RESULT 4

```
Q36579 ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Polyprotein.

Query Match 35.9%; Score 1597.5; DB 2; Length 3011;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 61; Gaps 15;

QY 146 TDNSPPVVPQSFQVAHLHAPTGSKSTKPAVAAGYKVLVLPNSVAATLGFAYMSK 205
Db 1211 TDNSPPVVPQSFQVAHLHAPTGSKSTKPAVAAGYKVLVLPNSVAATLGFAYMSK 1270
```

```
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT Transcribed RNA."
RL Science 277:570-574(1997).
DR EMBL; AF009606; AAB6324.1; -
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR PDB; 1N1L; X-ray; A/B=1017-1214.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR011410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_Ds_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327184 MW; E2E0EE809C63C1B9 CRC64;
```

```
QY 206 AHGIDNPRTGVRITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSILGIG 265
Db 1271 AHGVDNPRTGVRITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSILGIG 1330
QY 266 TVLDQAEATAGARLWLTATATPGSVTVPHNPINEVALSTTGEIPFYKKAIPLEVIKGRH 325
Db 1331 TVLDQAEATAGARLWLTATATPGSVTVSHNPINEVALSTTGEIPFYKKAIPLEVIKGRH 1390
QY 326 LIFCHSKKKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGVTGDFDS 365
Db 1391 LIFCHSKKKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGVTGDFDS 1450
QY 386 VIDCNTC----- 392
Db 1451 VIDCNCVTQTVDFSLDPTFTIETTLTPQDAVSRTQRRGTGRGKPGIYRFVAPGERPSG 1510
QY 393 ----- 392
Db 1511 MFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPLGPLVPCODHLEFWEVGFTGLTHIDA 1570
QY 393 ----- 392
Db 1571 HFLSOTKQSGENPPYLVAQATVCARAQAPPSPWDQMKCLIRLKPTLHGPTLLVRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEVTLTHPTIKYIMTCSADLEVTWTVLGVGLAALAAVCLSTGCVVIVIGRIVLSG 1690
QY 397 KPAILIPREVLYREFDEMESCQHLPIYIEQGMMLAEQFKOKALGL-----SRGKPAIVPD 452
Db 1691 KPAILIPREVLYREFDEMESCQHLPIYIEQGMMLAEQFKOKALGLLQTSRQAE-VITPA 1749
QY 453 KEVLYQOYDMEBCSQAAPYIEQAQVIAHQFKENVGLIDNDQV-----VT 499
Db 1750 VQTNWOKL-EVFWAKHWNFIISGIQYLAG-----LSTPGNPAIASLMFTAATVTSPLT 1802
QY 500 PDKEILYE-----AFDEMEECASKAALIEEGORMAELKSIQGLLG- 541
Db 1803 TGQTLFLNILGGVAAQAAPGAATAPVGAGLAGAATGSGVLGKVLVDILAGYGAGVAGA 1862
QY 542 -----ILRRHVGPGEAGVQWNNRLIAF 563
Db 1863 LVAFKIMSGEVPSTEDLVNLLPALISPGALWGVVCAAILRRHVGPGEAGVQWNNRLIAF 1922
QY 564 ASRGNHVSPTHYVPS----- 578
Db 1923 ASRGNHVSPTHYVPSDAAARVTAISSLVLTQLLRLHQLHWISSECTTPCSGSLWDIWD 1982
QY 579 ----- 578
Db 1983 WICEVLSDFKTLKAKLMPQLPGIPFVSCQGYGVWRGDMHTRCHCGAEITGHVKG 2042
QY 579 ----- 578
Db 2043 TMRIVGPTRCRNMWSGTFPINAVTTGCTPLPAPNYKFWALWRVSAEYVEIRRVGDPHYV 2102
QY 579 -----RSRFA-----QALPV 589
Db 2103 SGMTTNLIKPCQIPSPPEFTTELDGVRHLRFPAPCKPLLREEVSVFRVGLHEYVPGSQLPC 2162
QY 590 WARPDPY-----NPP----- 598
Db 2163 EPEPDVAVLTSMLTDPHSHTAAAGRLARGSPSSMASSASQLSAPSLKATCTANHDS 2222
QY 599 ---LVET---WKK-----PDYEPVPHG-----RSSRFA 622
Db 2223 DAELEIANLLWRQEMGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAAEILKSRFA 2282
QY 623 QALPVWARPDPYNPLVETWKKPDYEPVPHG 653
Db 2283 RALPVWARPDPYNPLVETWKKPDYEPVPHG 2313
```

```
RESULT 5
O36608 PRELIMINARY; PRT; 3011 AA.
AC O36608;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77.
RX MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR ENBL; AF011751; AAB67036.1; -.
DR PIR; A44150;
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P27958; 1HEI.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structeolysis and peptidolysis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR02521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept Ser Cys.
DR InterPro; IPR002518; Pept U39_HCV NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR SMART; SM00487; DEXDG; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327114 MW; 0B75B6B81CB5C198 CRC64;
```



```
Query Match      35.9%; Score 1597.5; DB 2; Length 3011;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

QY 146 TDNSSPPVPOSFOVAHLHAPTGGKSTKVPAAAYAAQYKVLVLPNSVAATLGFAYMSK 205
DB 1211 TDNSSPPVPOSFOVAHLHAPTGGKSTKVPAAAYAAQYKVLVLPNSVAATLGFAYMSK 1270

QY 206 AHGIDPNIRITGVRTITGSPITYTYGKFLADGCGSGAYDIIICDECHSDTATSILIG 265
DB 1271 AHGVDPNIRITGVRTITGSPITYTYGKFLADGCGSGAYDIIICDECHSDTATSILIG 1330

QY 266 TVLDQAEATAGARLVVLTATPPGSVTVPHNIEVALSTGEIPFYKKAIPLEVIKGRH 325
DB 1331 TVLDQAEATAGARLVVLTATPPGSVTVPHNIEVALSTGEIPFYKKAIPLEVIKGRH 1390

QY 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGVTGPDFS 385
DB 1391 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGVTGPDFS 1450

QY 386 VIDCNTC----- 392
DB 1451 VIDCNTCWTQTVDFSLDPTFTIETTLPODAVSRTQRRGCKPGIYRVAPGERPSG 1510

QY 393 ----- 392
DB 1511 MFDSSVLCYDAGCAWVELTPAETTVRLRAYMNTPLGVQDHLFEWGVFTGLTHIDA 1570

QY 393 ----- 392
DB 1571 HFLSQTKSGENFPYLVAYQATVCARAQAPPPSWDMQWKLIRLKPTLHGFTPLLRLGA 1630

QY 393 -----ACSG 396
DB 1631 VQNEVTLTHPTIKYIMTCSADLEWTVSTWLVGVVLAALAAAYCLSTGCVVIVGRIVLSG 1690

QY 397 KPAIIPDRELYRPFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
DB 1691 KPAIIPDRELYRPFDEMECSQHLPIYEQGMMLAEQFKQKALGLQATSRHAE-VITPA 1749

QY 453 KEVLYQYDEMECSQAAPYEQQAQVIAHQEKVGLINDQV-----VT 499
DB 1750 VQTWQKL-EVFWAKHMNFISGQYLAG-----LSTLPGNPAIASLMFAATAVTSPLT 1802

QY 500 PDKELIYE-----AFDEMECASKAALIEEGORMAEMLSKQGLIG- 541
DB 1803 TGTLLFNILGGVAAQAAPGAATAFVGAGLAGAATGSGVLGKVLVDILAGYAGVAGA 1862

QY 542 -----ILRRHVGPGEAGVQWNNRLIAF 563
DB 1863 LVAFKINSGEVPSSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWNNRLIAF 1922

QY 564 ASRGNHVSPTHYVS----- 578
DB 1923 ASRGNHVSPTHYVPSDAAARVTAISLSLTVTQLRLHOWISSECTTPCSCSWLRDIWD 1982

QY 579 ----- 578
DB 1983 WICEVLSDFKTLWAKMLPQLPGIPFVSCQRGVRVWRGDGIMHTRCHGAEITGHVQNG 2042

QY 579 ----- 578
DB 2043 TMRIVGPRTRCNMWSGTFPINAYTTGPTCTPLPAPNYKPALWRVSAEYVEIRRVGDHFVY 2102

QY 579 -----RSRRFA-----QALPV 589
DB 2103 SGMTDNLKCPQIPSPSEFTELDGVRLHRFPAPCKPLLRVEVSFRVGLHYEYPVGSQLP 2162

QY 590 WARPDPY-----NPP----- 598
DB 2163 EPEDPAVLVSLTDPSSHITAEAGRLRAGSPSPSMASASQLSAPSUKATCTANHDSP 2222
```

```
QY 599 ---LVET---WKK-----PDYEPVPVHG-----RSSRRFA 622
DB 2223 DAELTEANLWROBMGNITRVESENKVVLDSDPLVAEEDEREVSVPAILRKSRRFA 2282

QY 623 QALPVWARPDPYNNPPLVETWKKPDYEPVPVHG 653
DB 2283 RALPVWARPDPYNNPPLVETWKKPDYEPVPVHG 2313

RESULT 6
O36610 O36610 PRELIMINARY; PRT; 3011 AA.
AC O36610;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnaa.94.16.8738;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).
DR EMBL; AF011753; AAB67038.1; --
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P27958; 1HEI
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR02522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_Ds_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
```

DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RDRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA, 327222 MW, 293F91430A0D4067 CRC64;  
  
Query Match 35.9%; Score 1597.5; DB 2; Length 3011;  
Best Local Similarity 35.4%; Pred. No. 1.1e-89;  
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;  
  
QY 146 TDNSSPPVPOSQVQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNSVAATLFGAYMSK 205  
DB 1211 TDNSSPPVPOSQVQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNSVAATLFGAYMSK 1270  
  
QY 206 AHGIDPNIRGTVRTITGSPITYSTYTKFLADGGCGGAYDIIICDECHSTDATSILGIG 265  
DB 1271 AHGVDNIRGTVRTITGSPITYSTYTKFLADGGCGGAYDIIICDECHSTDATSILGIG 1330  
  
QY 266 TVLDQAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 325  
DB 1331 TVLDQAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 1390  
  
QY 326 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGVTGDFDS 385  
DB 1391 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGVTGDFDS 1450  
  
QY 386 VIDCNTC----- 392  
DB 1451 VIDCNTCVTVDFSLDPTFTIETTLPODAVSRTOGRGTGKPGIYRFVAPGRPSG 1510  
  
QY 393 ----- 392  
DB 1511 MFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGPVQCQDHLFEWEGFTGLTHIDA 1570  
  
QY 393 ----- 392  
DB 1571 HFLSQTQSGENFPYLVAQVATVCARAQAPPPSDQWQMKLIRLKPFLHGTPELLYRLGA 1630  
  
QY 393 -----ACSG 396  
DB 1631 VQNEVTLTHPTIKYIMTCMSADLEVTVTWLVGGVLAALAAVCLSTGCVIVGRIVLSG 1690  
  
QY 397 KPAILPDREVLYREFDEMESCQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452  
DB 1691 KPAILPDREVLYQEFDEMESCQHLPIYIEQGMMLAEQFKQKALGLLQQTASRHAETVTPA 1749  
  
QY 453 KEVLYQYDEMESCQAAPYIEQAQVTAHQFKEKVLGLINDQVV-----VT 499  
DB 1750 VQTNWQKL-EVFWAKHWNMFISGIQYLAG-----LSTLPGNPAIASLMAFTAAVTSPLT 1802  
  
QY 500 PDEKILYE-----AFDEMEECASKAALIEEGORMAELMKSITQGLLG- 541  
DB 1803 TGTLLFNILGGVAAQLAAPGRATAFVGAGLAGAAGTSGVLGKVLVDILAGYAGVAGA 1862  
  
QY 542 ----- 578  
DB 1863 LVAFKIMSGVSPSTEDLWLLPAILSPALVGVVCAAILRRHVGPGEAVQWNNRLIAF 1922  
  
QY 564 ASRGNHVSPTHYPS----- 578  
DB 1923 ASRGNHVSPTHYPSDAAARVTAISSLVTQLLRLHQMISSECTTPCSGSLWRDIWD 1982  
  
QY 579 ----- 578  
DB 1983 WICEVLSDFKTLWKALMPQLPGIPFVSCQGYRGVWRGDGIMHTRCHGCAEITGHVKNQ 2042  
  
QY 579 ----- 578  
DB 2043 TMRIVGPRTCRNWMSGTFPINAYTGPCTPLPAPNYKFAIWRVSAEYVEIRRVGDPHYV 2102

QY 579 -----RSRRFA-----QALPV 589  
DB 2103 SGMTTDNLKPCQIPSPPEFFTELDGVLHRFAPPCKPLLRREVSFRVGLHEYPVGSQLP 2162  
QY 590 WARPDP-----NPP----- 598  
DB 2163 EPEPDVAVLTSMITDPSHITABAAAGRRRLARGSPPSMASSASQSLKATCTANHDSP 2222  
QY 599 ---LVET---WKK-----PDYEPVVHG-----RSSRRFA 622  
DB 2223 DAELIEANLLMRQEMGNTRVESENKVVILDSFDPLVAEEDREVSVAEILKRRFA 2282  
QY 623 QALPVNWARDYNPPLVETWKKPDYEPVVHG 653  
DB 2283 RALPVNWARDYNPPLVETWKKPDYEPVVHG 2313  
  
RESULT 7  
Q9PMU9 PRELIMINARY; PRT; 3015 AA.  
AC Q9PMU9  
AT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Hepatitis C virus: an infectious molecular clone of a second major  
RT genotype (2a) and lack of viability of intertypic 1a and 2a  
RT chimeras.";  
RL Virology 262:250-263(1999).  
DR EMBL; AF177039; AAF01181.1; -;  
DR ENBL; AF177037; AAF01179.1; -;  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P27958; 1HEI.  
DR GO: GO:0016021; C:integral to membrane; IEA.  
DR GO: GO:0019028; C:viral capsid; IEA.  
DR GO: GO:0019031; F:ATP binding; IEA.  
DR GO: GO:0005524; F:ATP-dependent helicase activity; IEA.  
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO: GO:0019079; P:transcription; IEA.  
DR GO: GO:0019087; P:viral genome replication; IEA.  
DR InterPro: IPR000345; CytC\_heme\_BS.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR011545; DEAD/DEAH.  
DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS5a.  
DR InterPro: IPR002868; HCV\_Rdrp.  
DR InterPro: IPR002166; HCV\_Rdrp.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR004109; Peptidase\_S29.  
DR InterPro: IPR009003; Pept\_Ser\_Cys.  
DR InterPro: IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro: IPR002129; Pyridoxal\_deC.

```
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 35.9%; Score 1597.5; DB 2; Length 3015;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

QY 146 TDNSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLNPSVAATLGFAYMSK 205
DB 1215 TDNSPPAVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLNPSVAATLGFAYMSK 1274
QY 206 AHGIDPNRTGVRITTSPTSTYTGKFLADGGCGGAYDIIICDECHSDATSIILGIG 265
DB 1275 AHGVDPNRTGVRITTSPTSTYTGKFLADGGCGGAYDIIICDECHSDATSIILGIG 1334
QY 266 TVLQAEATAGARLVVLATATPPGTVTPHPNIEEVALSTTGEIPYGAIPLEVIKGRH 325
DB 1335 TVLQAEATAGARLVVLATATPPGTVTPHPNIEEVALSTTGEIPYGAIPLEVIKGRH 1394
QY 326 LIFCHSKKKDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYGDFDS 385
DB 1395 LIFCHSKKKDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYGDFDS 1454
QY 386 VIDNTC----- 392
DB 1455 VIDNTCTVTQTVDPSLDPTFTTTLTPQDAVSRQRTGRGKGIYRFVAPGERPSG 1514
QY 393 ----- 392
DB 1515 MFDSSVLCEYDAGCAWTELTPAETTVRLRAYMNTPGLPVCQDHLFEWGVFTGLTHIDA 1574
QY 393 ----- 392
DB 1575 HFLSQTKSGENFPYLVAQATVCARAQAPPSWDQMCKLIRLKPRTLHGPTPLLYRLGA 1634
QY 393 -----ACSG 396
DB 1635 VQNEVTLTHPIKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTGCWIVGRVILSG 1694
QY 397 KPAIIPDREVLVREFDEMECSCHLPYIEQGMLAEQFKQKALGH-----SRGKKAIVPD 452
DB 1695 KPAIIPDREVLVQEFDEMECSCHLPYIEQGMLAEQFKQKALGILLQTASRAAE-VITPA 1753
QY 453 KEVLYQQYDEMECSQAAPYIEQAQVIHQFKEVLGLINDQVV-----VT 499
DB 1754 VQTNNQKL-EVFWAKHMWNFISSIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLT 1806
QY 500 PKKEILYE-----AFDEMECSKAALIEGQMAEMKSKIOGLLG- 541
DB 1807 TCQTLLFNILGGWAAQLAAPGAATAFVAGLAGAAGISVGLGVLVDILAGYGAGVAGA 1866
QY 542 -----ILRRHVGGEGAVQMMNRLIAF 563
DB 1867 LVAFKIMSGEVSTEDLVNLLPAIILSPGALVGVVCAAILRRHVGGEGAVQMMNRLIAF 1926
QY 564 ASRGNHVSPTHYVPS----- 578
```

```
DB 1927 ASRGNHVSPTHYVPSDAAARVTAISSLTVTQLLRLLHQWISSECTTPCSGSWLRIWD 1986
QY 579 ----- 578
DB 1987 WICEVLSDFKTLAKAKLQPLPGIPFVSCQGRGVRGDMGIMHTRCHCGAEITGHVXNG 2046
QY 579 ----- 578
DB 2047 TMRIVGPRTCRNWMSGTFPINAYTTGPTCTPLPAPNYKFPALWRVSAEYVEIRRVGDFHYV 2106
QY 579 -----RSRRFA-----QALPV 589
DB 2107 SGMTDNLKPCQIPSPSEFFTELDGVRLLHRFAPCKPLLRREVSFRVGLHYPVCSQLPC 2166
QY 590 WARDPY-----NPP----- 598
DB 2167 EPEPDVAVLTSMLTDPSPHITAEAAAGRRRLARGSPSPMASSASQLSAPSLKATCTAHNDSP 2226
QY 599 ---LVET---WKK-----PDVEPPVHG-----RSSRRA 622
DB 2227 DAELIEANLLWRQEMGNITRVESENKVILDSFDPLVAEDEREVSVPAILRKSRRRA 2286
QY 623 QALPVMARPDYNPLVETWKKPDYEPVHG 653
DB 2287 RALPVMARPDYNPLVETWKKPDYEPVHG 2317

RESULT 8
Q9PMX5 PRELIMINARY; PRT; 3015 AA.
ID Q9PMX5;
AC Q9PMX5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras";
RL Virology 262:250-263(1999).
DR EMBL; AF177040; AAF01182.1; -.
DR EMBL; AF177038; AAF01180.1; -.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSP; P27958; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0003907; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR00345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
```

```
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR002129; Pyridoxal_deC.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BC1F190663A CRC64;

Query Match 35.9%; Score 1597.5; DB 2; Length 3015;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

Qy 146 TDNSPPVQSFQVHLHAPTSGKSTKVPAAVAAQGYKVLNPNVAATLGFAYMSK 205
Db 1215 TDNSPPVQSFQVHLHAPTSGKSTKVPAAVAAQGYKVLNPNVAATLGFAYMSK 1274

Qy 206 AHGIDPNIRGTGRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGIG 265
Db 1275 AHGVDPNIRGTGRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGIG 1334

Qy 266 TVLDAQETAGARLVLATATPPGSVTPPHNIEEVALSTTGEIPFYGKAIPLEVIKGRH 325
Db 1335 TVLDAQETAGARLVLATATPPGSVTPPHNIEEVALSTTGEIPFYGKAIPLEVIKGRH 1394

Qy 326 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTGTGTFDPS 385
Db 1395 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPTSGDVVVVSTDALMTGTFGDFDS 1454

Qy 386 VIDNCTC----- 392
Db 1455 VIDNCTCTQTVDPSLDPTFTIETTTLPQDAVSRQRTGRGKPGIYRFVAPGERPSG 1514

Qy 393 ----- 392
Db 1515 MFDSSVLCECYDAGCANYELTPAETTVRLRAYMNTPLPVCQDHLFEGVFTGLTHIDA 1574

Qy 393 ----- 392
Db 1575 HFLSQTKQSGENFPYVAYQATVCARAQAPPPSWDMWKCLIRLKLPTLHGPTLLYRLGA 1634

Qy 393 -----ACSG 396
Db 1635 VQNEVTLTHPTIKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRIVLSG 1694

Qy 397 KPAPIIPREVLRYREFDEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIYVD 452
Db 1695 KPAPIIPREVLRYQEFDEMECSQHLPIYIEQGMMLAEQFKQKALGLLTASRHAELVITPA 1753

Qy 453 KEVLVYQYDEMECSQAQPIYEQAOVTAHQPEKVLGLINDQVV-----VT 499
Db 1754 VQTNWQKL-EVFWAKHWMNFISGIQYLAG-----LSTLPGNPAIASLMAFTAAVTSPLT 1806
```

```
Qy 500 PDKEILYE-----AFDEMEECASKAALIEEGORMAEMLKSKIQTGLLG- 541
Db 1807 TGQTLLENILGCVAAQLAAPGAATAFVGAGLAGAALIGSVGLKVLVDILAGYAGVAGA 1866

Qy 542 -----ILRRHVGPGEAGAVQVMNRLIAF 563
Db 1867 LVAFKIMSGEVPSTEDLVNLLPALSPALVGVVCAAILRRHVGPGEAGAVQVMNRLIAF 1926

Qy 564 ASRGNHVSPTHYVPS----- 578
Db 1927 ASRGNHVSPTHYVPSDAAARVTAISSLVTQTLRLRLHQMVISSECTTPCGSMLRDIWD 1986

Qy 579 ----- 578
Db 1987 WICEVLSDFKTWLKAKLMPOLPGIPFVSCQGYGVWRGDGIMHTRCHCAEITGHVKNQ 2046

Qy 579 ----- 578
Db 2047 TMRIVGPRTCRNMWSGTGTFPINAVTTGCTPLPAPNVYKFWALRVSAEYVEIRRVGDFHYV 2106

Qy 579 -----RSRRFA-----QALPV 589
Db 2107 SGMTTNLIKPCQIPSPFETELDGVRLLHRFAPPCPLLRREVSFRVGLHEYPVGSQPLC 2166

Qy 590 WARPDP-----NPP----- 598
Db 2167 EPEPDVAVLTSMLTDPHSHTAAAGRRRLARGSPSMASASQLSAPSLKATCTANHDS 2226

Qy 599 ---LVET---WKK-----PYEPPVPHG-----RSSRRFA 622
Db 2227 DAELIEANLLWRQEMGNITRVESENKVVILDSFDPLVAEEDREVSVPAEILRKSRFA 2286

Qy 623 QALPVVARDPYNPLVETWKKPDYEPVPHG 653
Db 2287 RALPVVARDPYNPLVETWKKPDYEPVPHG 2317

RESULT 9
Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishihiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H., Tada F., Miyakawa Y., Mayumi M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
```

RA RN SEQUENCE FROM N.A.  
RX MEDLINE=94174722; PubMed=7510436;  
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,  
RA Inchauste G.;  
RT "Characterization and mapping of a B-cell immunogenic domain in  
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";  
RL Virology 200:246-255(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Okamoto H.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D10749; BAA01582.1; -;  
DR PIR: PS0326; PS0326.  
DR PIR: PS0327; PS0327.  
DR PIR: PS0328; PS0328.  
DR PIR: S40770; S40770.  
DR HSP; P26664; IHEI.  
DR GO: GO:0016021; C: integral to membrane; IEA.  
DR GO: GO:0019028; C: viral capsid; IEA.  
DR GO: GO:0019031; C: viral envelope; IEA.  
DR GO: GO:0005524; F: ATP binding; IEA.  
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.  
DR GO: GO:0003723; F: RNA binding; IEA.  
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
DR GO: GO:0005198; F: structural molecule activity; IEA.  
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
DR GO: GO:0006350; P: transcription; IEA.  
DR GO: GO:0019079; P: viral genome replication; IEA.  
DR GO: GO:0019087; P: viral transformation; IEA.  
DR InterPro: IPR000345; CytC\_heme\_BS.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR011545; DEAD/DEAH\_N.  
DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRP.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR004109; Peptidase\_S29.  
DR InterPro: IPR009003; Pept\_Ser\_Cys.  
DR InterPro: IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro: IPR007095; RNA\_pol\_DS\_FS.  
DR InterPro: IPR007094; RNA\_pol\_Psvir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; Helicase\_C; 1.  
DR Pfam: PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327114 MW; 97E9052C0250463B CRC64;  
  
Query Match 35.8%; Score 1596; DB 2; Length 3011;  
Best Local Similarity 34.6%; Pred. No. 1.3e-89;  
Matches 398; Conservative 36; Mismatches 93; Indels 624; Gaps 18;  
  
QY 115 HVTG-GAARTSTGL-----TSLEFPGASQNIQLITSDNSSPPVPSQFOV 160  
DB 1175 HWGIFRAAVCTRGVAKAVDFIPVESLETTMRSP-----VFTDSSPPAVPQSFQV 1225

QY 161 AHLHAPTSGSKSTKVPAAAYAAQYKVLNPSVAATLGFAYMSKAHGIIDPNIRTVRTI 220  
DB 1226 AHLHAPTSGSKSTKVPAAAYAAQYKVLNPSVAATLGFAYMSKAHGIIDPNIRTVRTI 1285  
QY 221 TTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATSILGIGTGLDQAEATAGARLVV 280  
DB 1286 TTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATSVLGLTGLDQAEATAGARLVV 1345  
QY 281 LATATPPGSVTVPHNTEEEVALSTTGEIPYKKAIPLEVIKGRHLIFCHSHKKKDELA 340  
DB 1346 LATATPPGSITVPHNTEEEVALSTTGEIPYKKAIPLEAIKGRHLIFCHSHKKKDELA 1405  
QY 341 KLVALGINAVAYRGLDVSIVTSGDVVVVATDALMTGTYGDDSDVDCNTC----- 392  
DB 1406 KLVALGVNAVAYRGLDVSIVTSGDVVVVATDALMTGTYGDDSDVDCNTC----- 1465  
QY 393 ----- 392  
DB 1466 LDPTFTIETTLTPODAVSRRTQRRCRTGRGKPGIYRFVAPGERPSGMPDSSILCECYDTC 1525  
QY 393 ----- 392  
DB 1526 AWYELTPAETTVRLRAYMNTPLPVCQDHLFEFMEGVFTGLTHIDAHFLSOTKOGGENPPY 1585  
QY 393 ----- 392  
DB 1586 LVAYQATVCARAQAPPSPDMMKCLIRLKLPTLHPTLLYRLGAVOGEVTLTHPVTKYI 1645  
QY 393 -----ACSGKPAIIPDREVLYREF 411  
DB 1646 MTCMSADLEVVTTWLVGGVLAALAAAYCLSTGCVVIVGRVLSGRPAIIPDREVLYREF 1705  
QY 412 DEMEECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIPVDKVELVYQYD-----E 462  
DB 1706 DEMEECSQHLPIYIEQGMMLAEQFKQKALGLLQTLASQAE-VIAPTVQTNMOKLEAFMAKH 1764  
QY 463 MEESQAAPIYIEQAQVIAHQFKEKVLGLINDQVVVTP---DKELIYE----- 507  
DB 1765 MNFISGIQVLAGLSTLPG--NPAAIASMATAAFTAVTSPLTTSQTLLFNILGQVVAQA 1822  
QY 508 -----AFDEMEECASKAALIEEGORMAEMLSKIQGLLG----- 541  
DB 1823 PGAATAFVSGSLAGAAVSGVLGRVLDILAGYAGVAGALVAFKINSGLPSTEDLVNL 1882  
QY 542 -----ILRRHVGPGEAGVQWMNRLIAFASRGNHVSPTHYVPS----- 578  
DB 1883 LPAILSPGALVGVVCAAILRRHVGPGEAGVQWMNRLIAFASRGNHVSPTHYVPSDAAA 1942  
QY 579 ----- 578  
DB 1943 RVTAISSLTVTQLRLRLHOWLSSESTTPCSCSWLRDIWDWICEVLSDFKTWLTKLMPH 2002  
QY 579 ----- 578  
DB 2003 LPGIPFVSCQHGKGVGRGDGIMHTRCHGAEITGHVKNGTMRIVGPKTCRNWMSGTPTPI 2062  
QY 579 ----- 578  
DB 2063 NAYTTGCTPLPAPNYTFALWRVSAEYVEIRRVGDPHYVTGWTDTNLKCKQVPSPEFF 2122  
QY 579 -----RSRRFA-----QALPVWARPDI----- 595  
DB 2123 TELDGLVRLHFPAPCKPLLREEVSFRVGLHDYPVGSQPLPCPEPEPDVAVLTSLMTPDSHIT 2182  
QY 596 -----NPP-----LVET---WKK----- 605  
DB 2183 AAAAGRRLARGSPSESSASSASQSLKATCTINHSDPAELIEANLLMRQEMGNIT 2242  
QY 606 -----PDYEPVPHG-----RSSRRFAQALPVWARPDPYNPPLVETWK 642  
DB 2243 RVESENKVVILDSFDPLVAEEDEREISVPAETLRKSRRTQALPIWARPDPYNPPLIETWK 2302  
QY 643 KPDYEPVPHG 653



```
QY 590 WARPDPY-----NRP-----598
Db 2163 EPEPDVAVLTSMITDPSHITAEBAGRLARGSPSPMASSASQSLKATCATANHDS 2222
QY 599 ---LVER---WKK-----PDYEPVHVG-----RSSRFA 622
Db 2223 DAELEIANLWROBMGNITRVSEKNKVLDSGDFPLVASEDEREVSVPAEILKRSRFA 2282
QY 623 QALPVWARPDPYNPLVETWKKPDYEPVHVG 653
Db 2283 PALPVWARPDPYNPLVETWKKPDYEPVHVG 2313

RESULT 11
Q6IX04
ID Q6IX04 PRELIMINARY; PRT; 2908 AA.
AC Q6IX04; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1a;
RA Brann T.W., Kottill S., Polis M., Imamichi T.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY615798; AAT44836.1; -.
DR HSP; P27958; 1A1V.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F: hydrolase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0008236; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0005198; F: serine-type peptidase activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39 HCV NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
```

```
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.
FT NON TER 2908 2908
SQ SEQUENCE 2908 AA; 315737 MW; BF5A4BC591498A4F CRC64;

Query Match 35.6%; Score 1586.5; DB 2; Length 2908;
Best Local Similarity 35.0%; Pred. No. 5e-89;
Matches 387; Conservative 32; Mismatches 86; Indels 601; Gaps 15;

QY 146 TDNSSPPVPSQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db 1211 TDNSSPPVPSQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270
QY 206 AHGIDPNIRTVGRTITITGSPITYTYGKFLADGSCSGAYDIIFICDECHSDTATSILIG 265
Db 1271 AHGVDPNIRTVGRTITITGSPITYTYGKFLADGSCSGAYDIIFICDECHSDTATSILIG 1330
QY 266 TVLDQAEATAGARLVLTATATPPGQSVTPHPNIEEVALSTGEIPFYGKATPLEVIKGRH 325
Db 1331 TVLDQAEATAGARLVLTATATPPGQSVTPHPNIEEVALSTGEIPFYGKATPLEVIKGRH 1390
QY 326 LIFCHSKKCDLAALKLVALGINAVYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 385
Db 1391 LIFCHSKKCDLAALKLVALGINAVYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 1450
QY 386 VIDCNTC-----392
Db 1451 VIDCNTCTQTVDFSLDPTTETITTLPODANVSRTQRRGTRGKPGYRFVAPGERPSG 1510
QY 393 -----392
Db 1511 MFDSSVLCVCYDAGCAWYELTPAETTVRLRAYMNTPLGVPQDHLFEWEGVFTGLTHIDA 1570
QY 393 -----392
Db 1571 HFLSQTQKSGENLPYLVAQATVCARAQAPPPSWDQMKLIRUKPTLHGTPPLLYRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEVTLTHPTVKYIMTCSADLEVVTSTWLVGVLAALAAAYCLSTGCVVIGRVLVS 1690
QY 397 KPAIIPREVLVREFDEMEBCSQHLPYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
Db 1691 KPAVIPREVLVREFDEMEBCSQHLPYIEQGMMLAEQFKQKALGLQLOATASQAE--VIAPA 1749
QY 453 KEVLYQQYD-----EMECCQAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEI 504
Db 1750 VQTNWQKLEAFWAKHMWNFIISGQYLAGLSTLPG--NPATASLMAFTAAVTSPLTTSOTL 1807
QY 505 LYE-----AFDEMEBCSKAALIEEGQMAEMLKSKIQGLLG-----541
Db 1808 LFNILGSWVAQAALAPGNAATAFVCGAGLAGAIGSVGLKVLVDILAGYGVAGALVAFK 1867
QY 542 -----ILRRHVGPGEQAVQMMNRLIAPASRGN 568
Db 1868 IMSGEPTTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEQAVQMMNRLIAPASRGN 1927
QY 569 HVSTHYVPS-----578
Db 1928 HVSTHYVPSDAAAARVTAISSLTVTXLLRLRHQWVSSECTTPCSGSNLKDINDWICEV 1987
QY 579 -----578
Db 1988 LSDFKTLWKALMPQLPGIIPFVSCQRYGVWGGIMHTRCHGAEITGHVKNGTMRIV 2047
QY 579 -----578
Db 2048 GPKTCRNWMSGTFPINAATTGCTPLPAPNYTFALWRVSAEYVEIRRVGDFHYVTGMTA 2107
QY 579 -----RSTRFA-----QALPVWARPDP 594
```



```
Db 2108 DNLKPCQVPSPFFTELDGVLRLHFRAPPCKPCLLRDEVSFRVGLHDYVPVGSQLPCEPEPD 2167
Qy 595 Y-----NPP-----LV 600
Db 2168 VAVLTSMLTDPHSHTAEARRRLARGSPPEASSASQSLRATCTTNNHSDPAELI 2227
Qy 601 ET---WKK-----PDYPPVVHG-----RSSRRFAQALPV 627
Db 2228 EANLLWRQMGGNITRVESENKVVILDSFDPLVAEEDEREVSAAEILRLKRRFTPALPI 2287
Qy 628 WARPDPNPPLVETWKPDYPPVVHG 653
Db 2288 WARPDPNPPLLEWKKPDYEPPTVHG 2313

RESULT 12
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC O9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RA Desai S.M., Devare S., Yamaguchi J.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF290978; AAGO2099.1; -.
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P26664; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR02522; HCV capsid.
DR InterPro; IPR02521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR02531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
```

```
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327108 MW; A6BECF5A3B3EE13F CRC64;

Query Match 35.5%; Score 1583.5; DB 2; Length 3011;
Best Local Similarity 35.0%; Pred. No. 8e-89;
Matches 387; Conservative 35; Mismatches 83; Indels 601; Gaps 15;

Qy 146 TONSPPVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db 1211 TONSPPVAVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270
Qy 206 AHGIDENIRTVRTITTTGSPITVSTYKFLADGGCSGGAYDIIICDECHSTDATSLIG 265
Db 1271 AHGVDPIRNTGVRTITTTGSPITVSTYKFLADGGCSGGAYDIIICDECHSTDATSLIG 1330
Qy 266 TVLDQAEATAGARLVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVIKGRH 325
Db 1331 TVLDQAEATAGARLVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVIKGRH 1390
Qy 326 LIFCHSKKCDLAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTFDFDS 385
Db 1391 LIFCHSKKCDLAALKVALGINAVAYRGLDVSIVPASGDVVVSTDALMTGTFDFDP 1450
Qy 386 VIDCNTC----- 392
Db 1451 VIDCNTCVTQTVDFSLDPTFTIETTLTLPQDANSRTORRGRTGKPGIYRFVAPGRPSG 1510
Qy 393 ----- 392
Db 1511 MFDSSVLCBYDAGCAWYELTPAETTVLRAYNMTPLPVCODHLEFWEQVFTGLTHIDA 1570
Qy 393 ----- 392
Db 1571 HFLSOTKSGENFPYLDVAYQATVCARAQAPPSPDWQMKLIRLKTPLHGPTELLRLGA 1630
Qy 393 -----ACSG 396
Db 1631 VQNEITLTHPVTKYIMTCHSANPEVVTSTWLVGGVLAALAAAYCLSTGCWIVIGRIVLSG 1690
Qy 397 KPAAIPDREVLYREFDEMEECSQLHPYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
Db 1691 KPAAIPDREVLYQEFDEMEECSQLHPYIEQGMMLAEQFKQKALGLLQOTASRQAE-VITPA 1749
Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVTAHQFKEKVLGLINDQVVVTP---DKEI 504
Db 1750 VQTNQKLEAFWAKHMMNFISGTQYLAGLSTLPG--NPAIASLMAFTAAVTSLTTSQTL 1807
Qy 505 LYX-----AFDEMEECASKAALIEEGORMAEMKSKQLGLG----- 541
Db 1808 LFNILGGWAAQLAAPGAATAFVAGLAGAAGTSGVLGKVLVDILAGYNGVAGALVAFK 1867
Qy 542 -----ILRRHVGPGEGAVQVMNRLIAFASRGN 568
Db 1868 IMSGEVPSTEDLVNLLPAILSPGALVGVVGVCAAILRRHVGPGEGAVQVMNRLIAFASRGN 1927
Qy 569 HVSPTHYVPS----- 578
Db 1928 HVSPTHYVPSDAAARVATILSNLTVTQLRLRHQWIGSECTTPCSGMLRDIWDICEV 1987
Qy 579 ----- 578
Db 1988 LSDFKTFLKAKLMPQLPGIPFYVSCQRGVRGWDGIMHTRCHCGAEITGHVKNGTMRIV 2047
Qy 579 ----- 578
```

Db 2048 GPRTCRNWSGTFFINAYTTGCTPLPAPNYKFPALMRVSAEYVEIRRVGDFHYVSGWTT 2107  
Qy 579 -----RSRFA-----QALPVMWARP 594  
Db 2108 DNLKPCQIIPSEFFETDGLVRLHFAPPCKPLLRVEESPRVGLHEYPVGSQLPCEPEPD 2167  
Qy 595 Y-----NPP-----LV 600  
Db 2168 VAVLTSLMTPSHITAEAGRLARGSPSSMASSASQLSAKATCTTNHSDPDAELI 2227  
Qy 601 ET---WKK-----PDYEPVVG-----RSRREAQALPV 627  
Db 2228 EANLLWRQMGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAILRKSQRFARALPV 2287  
Qy 628 WARDYNPPLVETWKKPDYEPVVG 653  
Db 2288 WARDYNPPLIETWKEPDYEPVVG 2313  
RESULT 13  
Q9DIT6 PRELIMINARY; PRT; 3011 AA.  
ID Q9DIT6  
AC Q9DIT6  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2004 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Genomic RNA for polyprotein gene.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=111103;  
RN 1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21014672; PubMed=11115058;  
RA Kumar U., Tsuchi T., Thomas H.C., Monjardino J.;  
RT "Sequence, expression and reconstitution of an HCV genome from a  
RT British isolate derived from a single blood donation.";  
RL J. Viral Hepat. 7:459-465(2000).  
DR EMBL; AJ278830; CAC03609.1; --  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P27958; 1A1V.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005199; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH N.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002868; HCV NS5a.  
DR InterPro; IPR002166; HCV RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF07652; Flavi\_DEAD; 1.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01540; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
KW PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT CHAIN 810 1026 non-structural protein 2.  
FT CHAIN 1027 1657 non-structural protein 3.  
FT CHAIN 1658 1711 non-structural protein 4a.  
FT CHAIN 2 191 core protein.  
FT CHAIN 1712 1972 non-structural protein 4b.  
FT CHAIN 1973 2420 non-structural protein 5a.  
FT CHAIN 2421 3011 non-structural protein 5b.  
FT CHAIN 192 383 envelop protein 1.  
FT CHAIN 384 809 envelop protein 2.  
SQ SEQUENCE 3011 AA; 327406 MW; 7B6264A74A5452D3 CRC64;  
Query Match 35.3%; Score 1574; DB 2; Length 3011;  
Best Local Similarity 33.2%; Pred. No. 3.1e-88;  
Matches 404; Conservative 39; Mismatches 121; Indels 654; Gaps 20;  
Qy 146 TDNSPPVVFQFVAHLHAPTSGSKSTKVPAAVAAQYKVLVLPNSVAATLGFAYMSK 205  
Db 1211 TDNSAPPAVVFQFVAHLHAPTSGSKSTKVPAAVAAQYKVLVLPNSVAATLGFAYMSK 1270  
Qy 206 AHGIDPNIRTVRTITGSPITYTYGKFLADGCGSGAYDIIICDECHSTDATSIIGIG 265  
Db 1271 AHGIDPNIRTVRTITGSPITYTYGKFLADGCGSGAYDIIICDECHSTDATSIIGIG 1330  
Qy 266 TVLDOAETAGARLVLTATATPPGTVVPHNIEEVALSTTGTGEPFYGKAIPLVETKGRH 325  
Db 1331 TVLDOAETAGARLVLTATATPPGTVVPHNIEEVALSTTGTGEPFYGKAIPLVETKGRH 1390  
Qy 326 LIFCHSKKKCDLAALKVALGINAVAYRGDLVSVIPTSGDVVVVATDALTMTGYTDFDS 385  
Db 1391 LIFCHSKKKCDLAALKVALGINAVAYRGDLVSVIPTSGDVVVVATDALTMTGYTDFDS 1450  
Qy 386 VIDNTC----- 392  
Db 1451 VIDNTCVTQTVDPSLDPTFTIETTLTPQDAVSRTRGRGKPGIYRFVAPGERSG 1510  
Qy 393 ----- 392  
Db 1511 MFDSAVLCECYDAGCANYELTPAETTVRLRAYMNTPGLPVCDHLEFWEVGTGLTHIDA 1570  
Qy 393 ----- 392  
Db 1571 HFLSQTOSGENLPYLVAQVATVCARAQAPPSNDQMKLIRLKLPTLHSTPLLYRLGA 1630  
Qy 393 -----ACSG 396  
Db 1631 VQNEVTLTHPTVKYIMTCMSADLEIVTSTWVLVGLVLAALAAAYCLSTGCWVIIVRVLISG 1690  
Qy 397 KPAIIPDREVLRYFDEMECSOHLPTYEOGWMIAEQFKOKALGL-----SRGKGAIVPD 452  
Db 1691 KPPIIPDQEVLYREFDEMECSOHLPTYEPGWMIAEQFKOKALALLQATASRAE-VIAPA 1749  
Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLINDNDQVVVTP---DKBI 504  
Db 1750 VQTNWQRLEAFWAKHWNFIIGIQVLAGLSTLPG--NPAIASLMAFTAATVTSPLTTSQTL 1807

```

Qy 505 LYE-----AFDEMEBCASKAALIEEGQRMALMSKIKGLLG----- 541
Db 1808 LFNILGCVAAQLAAPGAATAFVAGLAGAIGSVGLGKVLVDLAGVGAGVAGALVAFK 1867
Qy 542 -----ILRRHVGPGECAVQMMRLIAFASRGN 569
Db 1868 IMSGEVSTEDMWNLLPAILSPGALVGVVVCATILRRHVGPGECAVQMMRLIAFASRGN 1927
Qy 569 HVSPTHVUPS----- 578
Db 1928 HVSPTHVPSDAAARVAILSSLTVTQLRLRHQWSSSTTPCSCSWLRDIWDWICEV 1987
Qy 579 ----- 578
Db 1988 LSDFKTLWAKMLQPLGIPLVSCQGVGVQMGDGMVHTCHGCAITGHVKNGTWRIV 2047
Qy 579 ----- 578
Db 2048 GPKTCRNMSGTFPINAITYTTCPTPLPAPNYTFALMRVSAEYVEIRRVGDFHYVTGWT 2107
Qy 579 -----RSREA-----QALPVWARP 594
Db 2108 DNLRCPCQVSPFEFTLGDVLRHFAPPCKPLLRDEVSRVGLHDYVGSQLPCEPDP 2167
Qy 595 Y-----NPP-----LV 600
Db 2168 VAVVTSMLTDPSHITAEAGRLRAGSPSPVASSASQLSAPSLKATCTTNHSDPDAELI 2227
Qy 601 ET---WKK-----PDYEPVVHG-----RSRRFAQALPV 627
Db 2228 EANLWRQMGGNITRVESENKVVVLDSFDPLVAEEDEREVSVAEILTKSRRAEALAI 2287
Qy 628 WARDYNPPLVETWKDDYPPVHVGRKTKENTNRRPDQVKFG----- 671
Db 2288 WARDYNPPLLETWKDDYPPVHVHGCLP-----PPQSPVPSPRKRRTVVLTESTVST 2342
Qy 672 -----GGQIVGVGYLLPRRGLRLGLVLRKTSPIPKARPEGRKTWAQGYPPWPLY 721
Db 2343 ALAELATKSGSSSTSGI-----VGDSTTTPSPAPSVRSPDSDAECSSMP-PL 2392
Qy 722 GNKDRRSTGKSWGKPGYP 739
Db 2393 -----GEPGDP 2398

RESULT 14
POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270; DOI=10.1016/S0969-2126(98)00010-0;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -|- FUNCTION: Protease NS2 is responsible for the cleavage of NS2-NS3.
CC -|- FUNCTION: Protease NS3 is responsible for the cleavage of NS3-
CC NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B.
CC -|- FUNCTION: NS4A forms a complex with NS3 and is essential for the
CC activation of NS3.
CC -|- FUNCTION: NS5A seems to have a transcriptional activatory role.
CC -|- FUNCTION: NS5B is a RNA-dependent RNA polymerase that plays an
CC essential role in the virus replication.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -|- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins: E1
CC and E2. The nucleocapsid is a complex of protein C and mRNA.
CC -|- PTM: The structural proteins C, E1 and E2 are produced by
CC proteolytic processing by the host signal peptidases.
CC -|- SIMILARITY: Contains 1 peptidase S29 domain.
CC -|- SIMILARITY: Contains 1 peptidase U39 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA45534.1; -.
DR PIR; A36814; GNAVCH.
DR PDB; 1AIR; X-ray; A/B=1017-1214, C/D=1676-1698.
DR PDB; 1AIV; X-ray; A=1192-1667.
DR PDB; 1HEI; X-ray; -.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR004109; PeptIdase_S29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.

```

DR	Pfam; PF01506; HCV_NS5a; 1.	
DR	Pfam; PF00271; Helicase C; 1.	
DR	Pfam; PF00998; Viral RdRP; 1.	
DR	SMART; SM00487; DEXDC; 1.	
KW	3D-structure; ATP-binding; Coat protein; Core protein;	
KW	Envelope protein; Helicase; Hydrolase;	
KW	Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;	
KW	Serine protease; Transferrase; Transmembrane.	
FT	INIT_MET 1	
FT	CHAIN 1	191
FT	CHAIN 192	383
FT	CHAIN 384	746
FT	CHAIN 747	809
FT	CHAIN 810	1026
FT	CHAIN 1027	1671
FT	CHAIN 1658	1757
FT	CHAIN 1712	1972
FT	CHAIN 1973	2420
FT	CHAIN 2421	3011
FT	TRANSMEM 347	369
FT	ACT_SITE 1083	1083
FT	ACT_SITE 1107	1107
FT	ACT_SITE 1165	1165
FT	NP_BIND 1230	1237
FT	SITE 1316	1319
FT	CARBOHYD 196	196
FT	CARBOHYD 209	209
FT	CARBOHYD 234	234
FT	CARBOHYD 305	305
FT	CARBOHYD 417	417
FT	CARBOHYD 423	423
FT	CARBOHYD 430	430
FT	CARBOHYD 448	448
FT	CARBOHYD 476	476
FT	CARBOHYD 532	532
FT	CARBOHYD 540	540
FT	CARBOHYD 556	556
FT	CARBOHYD 576	576
FT	CARBOHYD 623	623
FT	CARBOHYD 645	645
FT	STRAND 1032	1035
FT	HELIX 1039	1048
FT	STRAND 1050	1050
FT	STRAND 1059	1063
FT	STRAND 1068	1074
FT	TURN 1075	1076
FT	STRAND 1077	1081
FT	HELIX 1082	1085
FT	TURN 1086	1087
FT	STRAND 1090	1092
FT	TURN 1093	1094
FT	STRAND 1095	1097
FT	STRAND 1101	1103
FT	TURN 1104	1107
FT	STRAND 1108	1112
FT	TURN 1115	1116
FT	STRAND 1120	1120
FT	STRAND 1122	1122
FT	STRAND 1129	1133
FT	TURN 1135	1136
FT	STRAND 1139	1144
FT	TURN 1147	1148
FT	STRAND 1149	1157
FT	HELIX 1158	1161
FT	TURN 1162	1163
FT	TURN 1165	1166
FT	STRAND 1168	1170
FT	TURN 1172	1173
FT	STRAND 1176	1186
FT	TURN 1187	1188
FT	STRAND 1189	1197
FT	HELIX 1198	1205

FT	STRAND 1680	1687
FT	STRAND 1693	1694
SO	SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;	
Query Match		
Best Local Similarity 35.0%; Score 1559.5; DB 1; Length 3011;		
Matches 388; Conservative 32; Mismatches 80; Indels 611; Gaps 15;		
Qy	146	TDNSSPPVPSQFOVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db	1211	TDNSSPPVPSQFOVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270
Qy	206	AHGIDPNIRGTITTTGSPITTYTGKFLADGGCGGAYDIIICDECHSDATSIILGIG 265
Db	1271	AHGVDNIRGTITTTGSPITTYTGKFLADGGCGGAYDIIICDECHSDATSIISIG 1330
Qy	266	TVLDQAETAGARLVVLTATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLVIKGRH 325
Db	1331	TVLDQAETAGARLVVLTATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLVIKGRH 1390
Qy	326	LIFCHSKKCKDELAALKVVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTFDPS 385
Db	1391	LIFCHSKKCKDELAALKVVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTFDPS 1450
Qy	386	VIDCNTC----- 392
Db	1451	VIDCNTCVTVDFSLDPTFTTETTLPODAVSTQRGRTRGKPKGIYRFVAFGERPSG 1510
Qy	393	----- 392
Db	1511	MFDSSVCECYDAGCANVELTPAETTVLRAYMNTPGLPVQCDHLGFWGVFTGLTHIDA 1570
Qy	393	----- 392
Db	1571	HFLSOTKSGENFPYLVAYQATVCARQAQPPSWDMQMKLIRLKLPTLHGPTLLYRLGA 1630
Qy	393	-----ACSG 396
Db	1631	VQNEVTLTHPTIKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTCCVIVIGRVLSG 1690
Qy	397	KPAIIPREVLRYREFDEMECSQHLPYIEQGMMLAEQFKQKALG-----SRGKPAIIPD 452
Db	1691	KPAIIPREVLRYQEFDEMECSQHLPYIEQGMMLAEQFKQKALGLLOTASRAE-VITPA 1749
Qy	453	KEVLYQQYDEMECSQAAPYIEQAQVIAHQPKVKVLGIDNDQV-----VT 499
Db	1750	VQTNWQKL-EVFWAKHMWNFISGQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLT 1802
Qy	500	PKSEILYE-----APDEMECSKAALIEEGQMAEMLKSKIOGLLG- 541
Db	1803	TGOTLLFNILGGVAAQLAFAATAFVAGLAGAALDSVGLGVLDILAGYGAGVAGA 1862
Qy	542	-----ILRRHVGPGEVAVQMMNRLIAF 563
Db	1863	LVAFKMSGEVPTSTEDLVNLLPAILSPGALAVGVVFASILRRRVGFGGAVQMMNRLIAF 1922
Qy	564	ASRGNHVSPTHYVPS----- 578
Db	1923	ASRGNHVSPTHYVPSDAAARVAILSSLTVTQLRLHQMISBCTTTPCSGSMURDIWD 1982
Qy	579	----- 578
Db	1983	WICEVLSDFTWLKAKLMPQLPGIPFVSCORGYRQVWRGDMHTRCHGAEITGHVKNK 2042
Qy	579	----- 578
Db	2043	TMRIVGPRCKNMMSGTFFINAYTTGPTCLPAPNYKFAALWRVSAEEYVEIRRVCDPHV 2102
Qy	579	-----RSRRFA-----QALPV 589
Db	2103	SGMTDNLKCPQCIQPSPEFFTELDGVRLLHRFAPPCKPLLRBEVSFRVGLHEYPVGSQPLC 2162
Qy	590	WAPDPY-----NPP----- 598

```

Db 2163 EPSPDVAVLSTMLTDPSSHITAEAGRLARGSPSPMASSASQLSAPSLKATCTANHDSF 2222
Q9DTE8
Qy 599 ---LVET---WKK-----PDYEPVPHG-----RSSRRFA 622
Db 2223 DAELIEANTLRQMGCGNITRSEKVNKVVLDSPDLVAEDEDESVPAEILKSRFA 2282
Qy 623 QALPVWARPDPYNPLVETWKKPDYEPVPHG 653
Db 2283 PALPVWARPDPYNPLVETWKKPDYEPVPHG 2313

RESULT 15
ID Q9DTE8 PRELIMINARY; PRT; 3010 AA.
AC Q9DTE8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Serum;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.",
RL HEPATOL. Res. 20:161-171(2001).
DR EMBL: AB049089; BAB18802.1; -.
DR PIR: A61196; A61196.
DR PIR: PQ0804; PQ0804.
DR PIR: PS0329; PS0329.
DR HSP; Q8JYS1; 1CWX.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019087; P: viral genome replication; IEA.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH N.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR004109; Peptidase_S29.
DR InterPro: IPR009003; Rept_Ser_Cys.
DR InterPro: IPR002518; Rept_U39_Hcv_NS2.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV capsid.1.
DR Pfam: PF01542; HCV Core.1.
DR Pfam: PF01539; HCV env.1.
DR Pfam: PF01560; HCV NS1.1.
DR Pfam: PF01538; HCV NS2.1.

```

```

DR Pfam: PF02907; HCV_NS3.1.
DR Pfam: PF01006; HCV_NS4a.1.
DR Pfam: PF01001; HCV_NS4b.1.
DR Pfam: PF01506; HCV_NS5a.1.
DR Pfam: PF00271; Helicase_C.1.
DR Pfam: PF00998; Viral_RdRp.1.
DR SMART; SMO0487; DEADG.1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3010 AA; 327068 MW; E26F4D669A836C80 CRC64;

Query Match 35.0%; Score 1559; DB 2; Length 3010;
Best Local Similarity 32.4%; Pred. No. 2.7e-87;
Matches 395; Conservative 44; Mismatches 119; Indels 662; Gaps 17;

Qy 48 VHEFGDNTAGCTSAGPHFNPLSTRGCNCSYIPQHITGHRMAWKLGSAAARTSGFVSL--F 105
Db 1142 VRRRGDNRGSLSPRPISYILKSGSGGPLLCPSGHVVG-----IFRAAVCTRGVAKAVEF 1195
Qy 106 APGAKQNETHVTGGAARTTSGLTSLFSPGASQNIQLITSTDNSSPPVVPQSQVAHLHA 165
Db 1196 IP-----VESMETWRSP-----VFTDNSSPPAVPQTQVAHLHA 1230
Qy 166 PTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHIDINIRGTGVTITGSP 225
Db 1231 PTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHIDINIRGTGVTITGAP 1290
Qy 226 ITYSTYTKFLADGCGSGGAYDIIICDECHSTDATSILGTGVLDDQATAGARLVLATAT 285
Db 1291 ITYSTYTKFLADGCGSGGAYDIIICDECHSTSTSTILGIGTGLDQATAGARLVLATAT 1350
Qy 286 PGGSVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGGHRLIFCHSKKKCDELAALKLVAL 345
Db 1351 PGGSVTPHPNIEEVALSNTGEIPFYGKAIPLEVIKGGHRLIFCHSKKKCDELAALKLAL 1410
Qy 346 GINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDSVIDCNTCA----- 393
Db 1411 GVNNAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTF 1470
Qy 394 ----- 393
Db 1471 TIETTTVPQDAVSRSORRGRTGRGGIYRFVTPGERPSGMFDSVLCBYDAGCAWYEL 1530
Qy 394 ----- 393
Db 1531 TPAETSVRLRAYLNTPLGLPVCQDHLFEWEGVFTGLTHIDAHFLSQTQKAGDHFPLVAYQ 1590
Qy 394 ----- 393
Db 1591 ATVCARAQAPPPSDQMWKCLTELKPTLHGPTLLYRLGAVQNEVTLTHPTKFIWTCMA 1650
Qy 394 -----CSGKPAIIPDREVLYREFDEMEE 416
Db 1651 ADLEIATSTWLVGGVLAALAAAYCLTTGSSVIVGRILSGRPAAVDPREVLYQEFDEMEE 1710
Qy 417 CSOHLPLYEQGMMLARQFKALGSLRGKGKPAIVDPKEVLYQQYDEMEECSQAAPYTE-- 474
Db 1711 CASHLPYIEQGMQLARQFKKALGLLQ-----TATKQEAAPVVEK 1753
Qy 475 ----QAQVIAH-----QFKEKVLGLIDNDQV-----VTPDKELY----- 506
Db 1754 WRGLEAFWAKHMWNFTISGIQYLAGLSTLPCNPALIASIMAFSTASITPLTQHTLLFNILG 1813
Qy 507 -----EAFDEMEECASKAALIEEGORMAEMLKSKIQGLLG----- 541
Db 1814 GWVAAQLAPPASAFVAGIAGAAVGSIGLGKVLVDILAGYAGVAGALVAFKVMGEM 1873
Qy 542 -----ILRRHVGPGEVAVQVMNRLIAFASRGNHVSPTH 574
Db 1874 PSTEDMVNLLPALFSPGALVVGVCVCAILRRHVGPGEVAVQVMNRLIAFASRGNHVSPTH 1933
Qy 575 YVP--SRSRRAQALP-----VWARPDYNPLVETWKK----- 605

```

Db	1934	YVPESDAAARVTQVLSSLTTIQLLRLHQMINEDCSTPCSGSLURDVMWICTVLTDFKT	1993
Qy	606	-----	605
Db	1994	WLQSKLLPRMPGVFFPFCQRCYKGVWRGDBGVMHTTCPCGADITGHVKNGSMRIVGPKTCS	2053
Qy	606	-----PDY-----	608
Db	2054	NTWHGTFPVNAYTTGCPCTPSPAPNYSKALMRVAAEYVEVTRVGDFHYVTGMTTDNVKCP	2113
Qy	609	-----EPPVV-----	613
Db	2114	CQVPAPFEFFTELDGVRHLRVAPACKPLLRDEVTFQVGLNQYPVGSQLPCEPEPDVTVITS	2173
Qy	614	-----HG-----	615
Db	2174	MLTDPSHITAEAAKERLARGCPSLASSASQLSAPSLKATCTTYHGSPPDADLIEANLLW	2233
Qy	616	-----RSSRRFAQALPVWARPDY	633
Db	2234	RQEMGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILKTRRRFPFPALPIWARPDY	2293
Qy	634	NPPLVETWKKDDYEPVWVG	653
Db	2294	NPPLIESWRKPDYVPPVWVG	2313

Search completed: November 7, 2005, 20:17:05  
Job time : 136.241 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 19:57:16 ; Search time 126.969 Seconds  
(without alignments)  
3347.654 Million cell updates/sec

Title: US-10-658-782-6

Perfect score: 5912

Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRTSGKSWGKPGYWP 1099

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003s:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5912	100.0	1099	5	AAU76378 HCV multi
2	5912	100.0	1099	6	ABG72262 HCV multi
3	5912	100.0	1099	8	ADL66809 HCV multi
4	4032	68.2	829	5	AAL18690 Multiple
5	4032	68.2	829	7	ADC06769 Chimeric
6	4032	68.2	829	8	ADL66807 HCV multi
7	3829.5	64.8	1021	2	AAW34481 HCV anti
8	3829.5	64.8	1021	2	AAW40039 Fusion pr
9	3829.5	64.8	1021	5	AAE22050 pSOD/c200
10	3050.5	51.6	841	2	AAE68547 HCV prote
11	3050.5	51.6	841	6	ABO27020 Hepatitis
12	3050.5	51.6	841	7	ADA07875 HCV prote
13	3047.5	51.5	841	2	AAW01701 HSD-HCV
14	3047.5	51.5	841	2	AAW46397 Amino aci
15	3047.5	51.5	841	2	AAW97609 Amino aci
16	2042.5	51.5	840	2	AAE14349 HCV prote
17	2909.5	49.2	2261	1	AAE90164 Peptide e
18	2909.5	49.2	2436	1	AAE92050 HCV prote
19	2909.5	49.2	2436	1	AAE90288 Peptide e
20	2909.5	49.2	2772	3	AAE18540 Protein e
21	2909.5	49.2	2772	8	ADN35976 HCV cDNA
22	2909.5	49.2	2955	2	AAV14975 Amino aci
23	2909.5	49.2	2955	3	AAE18541 Polypeptide
24	2909.5	49.2	2955	8	ADN35978 HCV cDNA
25	2909.5	49.2	3011	2	AAE90931 Hepatitis

#### ALIGNMENTS

##### RESULT 1

AAU76378  
ID AAU76378 standard; protein; 1099 AA.

AC AAU76378;

DT 08-MAY-2002 (first entry)

DE HCV multiple epitope fusion antigen (MEFA) 7.1 protein sequence.

KW Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;  
KW non-structural protein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200196870-A2.

XX 20-DEC-2001.

XX 14-JUN-2001; 2001WO-US019156.

XX 15-JUN-2000; 2000US-0212082P.

XX 02-APR-2001; 2001US-0280811P.

XX 02-APR-2001; 2001US-0280867P.

PA (CHIR ) CHIRON CORP.

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;

DR WPI: 2002-090228/12.

XX N-PSDB; ABK15345.

PT Immunoassay solid support, useful for detecting hepatitis C virus  
infection in biological sample, comprises HCV NS3/4a conformational  
epitope and multiple epitope fusion antigen bound to the support.

XX Claim 5; Fig 5; 92pp; English.

XX The present invention relates to a new immunoassay solid support  
consisting essentially of at least one hepatitis C virus (HCV) NS3/4a  
conformational epitope and a multiple epitope fusion antigen (MEFA),  
bound to the support. The NS3/4a conformational epitope and/or MEFA  
reacts specifically with anti-HCV antibodies present in a biological  
sample from an HCV-infected individual. The immunoassay of the invention  
is useful for detecting hepatitis C virus infection in a biological

AAW34480 HCV polyp  
AAW40038 HCV polyp  
AAE22049 Hepatitis  
ADL22107 Hepatitis  
ADR29357 Hepatitis  
Aap92047 HCV prote  
Aar08123 Hepatitis  
Aar21519 Compiled  
AAU84597 HCV polyp  
AAE25135 HCV polyp  
Aap92041 Hepatitis  
Aar24440 Composite  
Aar34009 HCV-1 pol  
Aar31621 Hepatitis  
Aap90158 Protein s  
Aar70230 Composite  
Aar28582 HCV amino  
Aar08124 Hepatitis  
AAE22052 Hepatitis  
ADR38451 Hepatitis

26 2909.5 49.2 3011 2 AAW34480  
27 2909.5 49.2 3011 2 AAW40038  
28 2909.5 49.2 3011 5 AAE22049  
29 2909.5 49.2 3011 8 ADL22107  
30 2909.5 49.2 3011 8 ADR29357  
31 2906.5 49.2 2301 1 Aap92047  
32 2906.5 49.2 2301 1 Aar08123  
33 2900.5 49.1 3011 2 Aar21519  
34 2900.5 49.1 3011 5 AAU84597  
35 2897.5 49.0 2435 2 AAR25135  
36 2896.5 49.0 1766 1 Aap92041  
37 2895.5 49.0 2894 2 AAR24440  
38 2893.5 48.9 2816 2 AAR34009  
39 2891.5 48.9 3011 2 AAR31621  
40 2890.5 48.9 1786 1 AAP90158  
41 2890.5 48.9 2894 2 AAR70230  
42 2883.5 48.8 2436 2 AAR28582  
43 2880.5 48.7 2955 2 AAR08124  
44 2877.5 48.7 3011 5 AAE22052  
45 2875 48.6 1986 8 ADR38451



FT polypeptide C100"  
FT 884..919  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT 922..957  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT 958..1028  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
FT 1029..1099  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
XX US2002146685-A1.  
XX  
XX 10-OCT-2002.  
XX  
XX 14-JUN-2001; 2001US-00881654.  
XX  
XX 15-JUN-2000; 2000US-0212082P.  
XX 02-APR-2001; 2001US-0280811P.  
XX 02-APR-2001; 2001US-0280867P.  
XX  
XX (CHIE/) CHIEN D Y.  
XX (ARCA/) ARCANDEL P.  
XX (TAND/) TANDESKE L.  
XX (GEOR/) GEORGE-NASCIMENTO C.  
XX (COIT/) COIT D.  
XX (MED/) MEDINA-SELBY A.  
XX  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
XX Medina-Selby A;  
XX  
XX WPI; 2003-147573/14.  
XX N-PSDB; ABX14411.  
XX  
XX Immunocassay solid support for detecting Hepatitis C Virus infection in  
XX biological samples, comprises Hepatitis C Virus conformational epitope  
XX and multiple epitope fusion antigen.  
XX  
XX Claim 25; Fig 5A-5F; 45pp; English.  
XX  
XX The present invention relates to immunocassays comprising Hepatitis C  
XX Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
XX antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
XX multiple epitope fusion antigen react with anti-HCV antibodies present in  
XX a biological sample from an HCV-infected individual. The immunocassays and  
XX methods of the invention are useful for detecting HCV infection in a  
XX biological sample. The inventive immunocassay solid support provides a  
XX sensitive and reliable method for detecting early HCV seroconversion. The  
XX assays can detect HCV infection caused by any six known genotypes of HCV.  
XX The use of the multiple epitope fusion proteins decreases masking  
XX problems, improves sensitivity in detecting antibodies by allowing a  
XX greater number of epitopes on a unit area of substrate, and improves  
XX selectivity. The present sequence represents HCV multiple epitope fusion  
XX antigen 7.1 (MEFA 7.1), a mutant HCV polypeptide derived from various  
XX regions of HCV type 1, 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide  
XX sequences  
XX  
XX Sequence 1099 AA;  
XX  
XX Query Match 100.0%; Score 5912; DB 6; Length 1099;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFVHVEFDNTAGCTS 60  
XX  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFVHVEFDNTAGCTS 60  
XX  
XX 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLSDHCHCIIGRTLIV 120

RESULT 3



DE Multiple epitope fusion antigen (MEFA) 12 protein.  
XX Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;  
KW HCV infection; MEFA 12 protein.  
XX Unidentified.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 315 /note= "Encoded by ATG"  
FT FT  
FT Misc-difference 645 /note= "Encoded by GAG"  
FT FT  
XX WO200196875-A2.  
XX 20-DEC-2001.  
XX 14-JUN-2001; 2001WO-US019369.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIR ) CHIRON CORP.  
FA Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX  
XX WPI; 2002-179522/23.  
DR N-PSDB; AAD29796.  
XX  
XX Immunassay solid support useful for detecting hepatitis C virus  
PT infection in a biological sample, comprises at least one of HCV anti-core  
PT antibody and HCV NS3/4a epitope, bound to the support.  
XX  
XX Disclosure; Fig 7; 87pp; English.  
XX  
XX The present invention relates to hepatitis C virus (HCV) core antigen and  
CC NS (nonstructural) 3/4a antibody combination assay that can detect both  
CC HCV antigens and antibodies present in a sample using a single solid  
CC matrix as well as immunoassay solid supports for use in the assay. The  
CC solid support is useful for detecting HCV infection in a biological  
CC sample. The present sequence is MEFA (multiple epitope fusion antigen) 12  
CC protein. This sequence is used in the exemplification of the invention  
XX  
XX Sequence 829 AA;  
Query Match 68.2%; Score 4032; DB 5; Length 829;  
Best Local Similarity 69.8%; Pred. No. 2.2e-275;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
QY 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTGLGHGFVHEFGDNTAGCTS 60  
DB 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTGLGHGFVHEFGDNTAGCTS 60  
QY 61 AGPHNPLSRKHGGPKBERHVGDLGNVTADKGVADVSIEDSVLSLSDGHCIIIGRTLIV 120  
DB 61 AGPHNPLSTR-----  
QY 121 HEKADDLKGKNEESTKTGNAGSRLACGVIGIAQNLNSGCNCSYIPGHITGHRMAWKLS 180  
DB 72 -----GCNCSYIPGHITGHRMAWKLS 93  
QY 181 AARTTSGFVSLPAPKAKONETHVTGGAARTTSGLTSLFSPGASONIQIIVDFIPVENLE 240  
DB 94 AARTTSGFVSLPAPKAKONETHVTGGAARTTSGLTSLFSPGASONIQILTS----- 145  
QY 241 TTMRSPVFTDSSPPVWPQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAATL 300  
DB 146 -----TDNSSPPVWPQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLVLPNSVAATL 197  
QY 301 GFGAYMSKAHGIDPNIRTVRTITTTGSPITYSTYVGKFLADGCGSGGAYDIIICDECHSTD 360  
|||||

Db 198 GFGAYMSKAHGIDPNIRTVRTITTTGSPITYSTYVGKFLADGCGSGGAYDIIICDECHSTD 257  
QY 361 ATSLIGTGTVLDOAETAGARLVVLTATPPGCVTVPHNIEEVALSTGEIPFGYKAIP 420  
DB 258 ATSLIGTGTVLDOAETAGARLVVLTATPPGCVTVPHNIEEVALSTGEIPFGYKAIP 317  
QY 421 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 480  
DB 318 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 377  
QY 481 GYTGFDSVIDQNTCVTQTVDFTTITITLPODAVSRTQRRGRTGRGKPGYRPFV 540  
DB 378 GYTGFDSVIDQNTC----- 392  
QY 541 APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTFGLPVCQDHLSEFWGVF 600  
DB 393 ----- 392  
QY 601 TGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMWKCLIRLKLPHGPT 660  
DB 393 ----- 392  
QY 661 PLYRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSSACSGKPAIIPDREVLRYBDEMEE 720  
DB 393 -----ACSGKPAIIPDREVLRYBDEMEE 416  
QY 721 CSQHLPIYEOGMMLAEQFKKALGLSRGKPAIIPDREVLRYBDEMEECSQAAPYIRQA 780  
DB 417 CSQHLPIYEOGMMLAEQFKKALGLSRGKPAIIPDREVLRYBDEMEECSQAAPYIRQA 476  
QY 781 QVIAHQPKVKVGLIINDQVVTDPKELLYEAFDEMEECASKAALIEEGORMAELKSKI 840  
DB 477 QVIAHQPKVKVGLIINDQVVTDPKELLYEAFDEMEECASKAALIEEGORMAELKSKI 536  
QY 841 QGLLGILRRHVGPEGAVQWNRLLAFASRGNHVSPTHYVPSRERRFAQALPVWARPDPY 900  
DB 537 QGLLGILRRHVGPEGAVQWNRLLAFASRGNHVSPTHYVPSRERRFAQALPVWARPDPY 596  
QY 901 PPLVETWKKPDYEPVHGRSSRRFAQALPVWARPDPYVETWKKPDYEPVHGRKT 960  
DB 597 PPLVETWKKPDYEPVHGRSSRRFAQALPVWARPDPYVETWKKPDYEPVHGRKT 656  
QY 961 KRNTNRPPQDVKFGGGQIVG-----RRGP-----PIPKARRPEGRRTWAQPGY 1003  
DB 657 KRNTNRPPQDVKFGGGQIVG-----RRGP-----PIPKARRPEGRRTWAQPGY 716  
QY 1004 PWPLYGNKDRRSTGKSGKPGYWPWRKTKRNTNRPPQDVKFGGGQIVG-----RRGP- 1056  
DB 717 PWPLYGNKDRRSTGKSGKPGYWPWRKTKRNTNRPPQDVKFGGGQIVG-----RRGP- 776  
QY 1057 -----PIPKARRPEGRRTWAQPGYWPPLYGNKDRRSTGKSGKPGYWP 1099  
DB 777 LGVLATRKTSIPKARRPEGRRTWAQPGYWPPLYGNKDRRSTGKSGKPGYWP 829

RESULT 5  
ADC06769  
ID ADC06769 standard; protein; 829 AA.  
XX  
AC ADC06769;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Chimeric multiple epitope fusion antigen 12 protein.  
KW immunoassay solid support; HCV; NS3/4a; non-structural;  
KW non-A, non-B hepatitis; NANB; multiple epitope fusion antigen 12; MEFA12;  
KW chimeric.  
OS Chimeric.  
OS Synthetic.  
OS Unidentified.  
OS Hepatitis C virus.

OS Homo sapiens.  
XX US2002192639-A1.  
XX 19-DEC-2002.  
XX 14-JUN-2001; 2001US-00881239.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIE//) CHEN D Y.  
PA (ARCA//) ARCANDEL P.  
PA (TAND//) TANDESKE L.  
PA (GEOR//) GEORGE-NASCIMENTO C.  
PA (COIT//) COIT D.  
PA (MEDI//) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX WPI; 2003-644609/61.  
DR N-PSDB; ADC06770.  
XX  
XX Immunoassay solid support for detecting hepatitis C virus infection in  
PT biological samples, comprises a hepatitis C virus anti-core antibody and  
PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
PT antibody.  
XX  
XX Claim 45; Fig 7; 40pp; English.  
PS  
XX The invention relates to a novel immunoassay solid support comprising at  
CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.  
CC The system of the invention may be useful for detecting HCV infection in  
CC a biological sample and for treating or detecting non-A, non-B hepatitis  
CC (NANB hepatitis). The current sequence is that of the chimeric multiple  
CC epitope fusion antigen 12 (MEFA12) protein of the invention.  
XX  
XX Sequence 829 AA;  
PS  
Query Match 68.2%; Score 4032; DB 7; Length 829;  
Best Local Similarity 69.8%; Pred. No. 2.2e-275;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
61 AGPHFNPLSRKHGGPKDEERHVGDLGNVTADKGVADVSIEDSVLSLGDHCHIIIGRTLTV 120  
61 AGPHFNPLSTR-----  
121 HEKADDLKGGNEESTKTGNAGSLACGVIGIAQNLSGNCISYPGHITGHRMAWKLS 180  
72 -----GCNCSYPGHITGHRMAWKLS 93  
181 AARTTSFVSLFAPGAKONETHVTGGAAARTTSLTSLFSPGASQNTQLIVDFIPVENLE 240  
94 AARTTSFVSLFAPGAKONETHVTGGAAARTTSLTSLFSPGASQNTQLITS----- 145  
241 TTRWSPVFTNNSPPVQSFQVLAHLHAPTCSGKSTKVPAAQAQGVKVLNLPNSVAATL 300  
146 -----TNSPPVQSFQVLAHLHAPTCSGKSTKVPAAQAQGVKVLNLPNSVAATL 197  
301 GFAYMSKAGIDNIRTYRTITTSPIITYSTYTKFLADGGCGSGAYDIIICDECHSTD 360  
198 GFAYMSKAGIDNIRTYRTITTSPIITYSTYTKFLADGGCGSGAYDIIICDECHSTD 257  
361 ATSLIGTGLVLDQAGTLAGRLVLTATPPGCVTVTPHPNIEEVALSTTGEPFYGKAIP 420  
258 ATSLIGTGLVLDQAGTLAGRLVLTATPPGCVTVTPHPNIEEVALSTTGEPFYGKAIP 317

QY 421 EVIKGGRHLIFCHSKKKCDELAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 480  
DB 318 EVIKGGRHLIFCHSKKKCDELAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 377  
QY 481 GYTGFDSVIDCNTCVTQTVDFSLDPTFTTITLPODAVSRQRRGRGRGKPGYRFV 540  
DB 378 GYTGFDSVIDCNTC----- 392  
QY 541 APGERPSGPFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVGF 600  
DB 393 ----- 392  
QY 601 TGLTHIDAHFLSOTKSGENLPYLVAQATVCARAQAPPSPDQMWKCLIRLKLPTLHGPT 660  
DB 393 ----- 392  
QY 661 PLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSACSGKPAIIPDREVLYREFDEMEE 720  
DB 393 -----ACSGKPAIIPDREVLYREFDEMEE 416  
QY 721 CSQHLPIEQGMMLABQFKQKALGLSRGKPAIVDPKEVLYQOYDEMEECQAAPYIEQA 780  
DB 417 CSQHLPIEQGMMLABQFKQKALGLSRGKPAIVDPKEVLYQOYDEMEECQAAPYIEQA 476  
QY 781 QVIAHOFKEKVLGLINDOVVTPDKIILYEAFADEMEECASKAALIEEGORMAEMLSKI 840  
DB 477 QVIAHOFKEKVLGLINDOVVTPDKIILYEAFADEMEECASKAALIEEGORMAEMLSKI 536  
QY 841 QGLLILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPYN 900  
DB 537 QGLLILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPYN 596  
QY 901 PPLVETWKKPDYPPVYVHGRSSRRFAQALPVWARPYNPPLVETWKKPDYPPVYVHGRKT 960  
DB 597 PPLVETWKKPDYPPVYVHGRSSRRFAQALPVWARPYNPPLVETWKKPDYPPVYVHGRKT 656  
QY 961 KRNTNRRPQDVKPPGGQIVG-----RRGP-----PIPKARRPEGRWTWQPGY 1003  
DB 657 KRNTNRRPQDVKPPGGQIVG-----RRGP-----PIPKARRPEGRWTWQPGY 716  
QY 1004 PWPLYGNKORRSTGKSWGKPGYPPWPRKTXTNTNRRPQDVKPPGGQIVG-----RRGP- 1056  
DB 717 PWPLYGNKORRSTGKSWGKPGYPPWPRKTXTNTNRRPQDVKPPGGQIVGVLPRRGR 776  
QY 1057 -----PIPKARRPEGRWTWQPGYPPWPLYGNKORRSTGKSWGKPGYPPW 1099  
DB 777 LGVLATRKTSPIPKARRPEGRWTWQPGYPPWPLYGNKORRSTGKSWGKPGYPPW 829  
RESULT 6  
ADL66807  
ID ADL66807 standard; protein; 829 AA.  
XX  
XX ADL66807;  
XX AC AC  
XX XX  
DT 03-JUN-2004 (first entry)  
XX  
DE HCV multiple epitope fusion antigen 12 (MEFA 12) polypeptide.  
XX  
XX HCV; MEFA 12; HCV antigen; HCV polyprotein;  
KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection;  
KW multiple epitope fusion antigen 12.  
XX  
XX Hepatitis C virus.  
XX  
XX WO2004021871-A2.  
XX PN  
XX 18-MAR-2004.  
XX PD  
XX 08-SEP-2003; 2003WO-US028071.  
XX PF  
XX 09-SEP-2002; 2002US-0409515P.  
XX PR

XX (CHIR ) CHIRON CORP.  
 PA Arcangel P, Chien D;  
 FI WPI; 2004-248333/23.  
 XX DR N-PSDB; ADL66806.  
 DR  
 XX Detecting hepatitis C virus (HCV) infection in a biological sample by  
 PT detecting complexes formed between the HCV antibody and the antigens from  
 PT the first region of the HCV polypeptide and the multiple epitope fusion  
 PT antigen (MEFA).  
 XX  
 XX Claim 14; SEQ ID NO 4; 93pp; English.  
 PS  
 XX The invention relates to a method of detecting hepatitis C virus (HCV)  
 CC infection in a biological sample. The method comprises providing an  
 CC immunassay solid support comprising HCV antigens bound to it, where the  
 CC HCV antigens comprise one or more isolated antigens form a first region  
 CC of the HCV polypeptide, combining a biological sample with the solid  
 CC support under conditions that allow HCV antibodies, when present in the  
 CC biological sample, to bind to the one or more HCV antigens, adding to the  
 CC solid support a detectably labelled HCV multiple epitope fusion antigen  
 CC (MEFA), where the labelled MEFA comprises at least one epitope from the  
 CC same region of the HCV polypeptide as the one or more isolated antigens,  
 CC where the MEFA binds to the bound HCV antibody, and detecting complexes  
 CC formed between the HCV antibody and the one or more antigens from the  
 CC first region of the HCV polypeptide and the MEFA, if any, as an  
 CC indication of HCV infection in the biological sample. The method is  
 CC useful for detecting hepatitis C virus (HCV) infection in a biological  
 CC sample. This sequence represents the MEFA 12 polypeptide used in the  
 CC scope of the invention.  
 XX  
 XX Sequence 829 AA;  
 SQ  
 Query Match 68.2%; Score 4032; DB 8; Length 829;  
 Best Local Similarity 69.8%; Pred. NO. 2.2e-275;  
 Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
 QY 1 MATKAVCVLKGDPVQGIINFEQESNGPVKWSIKGLTEGLHGFVHFEGDNTAGCTS 60  
 DB 1 MATKAVCVLKGDPVQGIINFEQESNGPVKWSIKGLTEGLHGFVHFEGDNTAGCTS 60  
 QY 61 AGPHENPLSRKHGGPKDEERHVDLGNVTADKGDVADVSDIEDSVLSGDHCHIIIGRTLVV 120  
 DB 61 AGPHENPLSTR----- 71  
 QY 121 HEKADDLGKGGNEESTKTGNAGSLACGVIGIAQNLSGNCISIYPGHITGHRMAWKLS 180  
 DB 72 -----GCNCSIYPGHITGHRMAWKLS 93  
 QY 181 AARTTSGFVSLPAPGAKQNEHTVTCGAARTTSGLTSLFSPGASONIOLIVDFIPVENLE 240  
 DB 94 AARTTSGFVSLPAPGAKQNEHTVTCGAARTTSGLTSLFSPGASONIOLITS----- 145  
 QY 241 TTMRSFVFTDNSSPPVPPQSFQVHLHAPTSGKSTKVPAAAGYKVLVLPNSVAATL 300  
 DB 146 -----TDNSSPPVPPQSFQVHLHAPTSGKSTKVPAAAGYKVLVLPNSVAATL 197  
 QY 301 GFAYMSKAHGIDPNIRTVRTITTTGSPITYSTYTGKFLADGCGSGAGYDIIICDECHSTD 360  
 DB 198 GFAYMSKAHGIDPNIRTVRTITTTGSPITYSTYTGKFLADGCGSGAGYDIIICDECHSTD 257  
 QY 361 ATSLIGIGVLDQATAGARLVVLTATPPGVSVPVPHNIEEVALSTGTEIPFYKAIPL 420  
 DB 258 ATSLIGIGVLDQATAGARLVVLTATPPGVSVPVPHNIEEVALSTGTEIPFYKAIPL 317  
 QY 421 EVIKGRHLIFCHSKKKCKDELAAKLVAGINAVAYYRGLDVSVIPTSGDVVVVATDALMT 480  
 DB 318 EVIKGRHLIFCHSKKKCKDELAAKLVAGINAVAYYRGLDVSVIPTSGDVVVVATDALMT 377  
 QY 481 GYTGFDSVIDNCNTCVTQTQVDFSLDFTFTTITLTPQDAVSRTQRRGRTGKPGIYRFV 540  
 |||

Db 378 GYTGFDSVIDNCNTC----- 392  
 QY 541 APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPEGLPVCQDHLSEWEGVF 600  
 Db 393 ----- 392  
 QY 601 TGLTHIDAHFLSQTQSGENLPYLVAQATVCABAQAPPSPSQDQMWKCLIRLKTPLHGPT 660  
 Db 393 ----- 392  
 QY 661 PLLVRLGAVQNEITLTHPTVKYIMTCSADLEVVTSACSGKPAIPDREVLYREFDEEMEE 720  
 |||-----ACSGKPAIPDREVLYREFDEEMEE 416  
 Db 721 CSQHLPIEQGMMLAEQFKQKALGLSRGGKPAIPVDKEVLYQQYDEMEECQAAPYIEQA 780  
 Db 417 CSQHLPIEQGMMLAEQFKQKALGLSRGGKPAIPVDKEVLYQQYDEMEECQAAPYIEQA 476  
 QY 781 QVIAHQFKEKVLGLINDQVVVTPDKELIYEADEFMEECASKAALIEEQORMAEMLKSKI 840  
 Db 477 QVIAHQFKEKVLGLINDQVVVTPDKELIYEADEFMEECASKAALIEEQORMAEMLKSKI 536  
 QY 841 QGLLGILRRHVGPCEGAVQWNRLLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPYN 900  
 Db 537 QGLLGILRRHVGPCEGAVQWNRLLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPYN 596  
 QY 901 PPLVETWKKPDYEPVHVHGRSSRRFAQALPVWARPDPYNPLVETWKKPDYEPVHVHGRKT 960  
 Db 597 PPLVETWKKPDYEPVHVHGRSSRRFAQALPVWARPDPYNPLVETWKKPDYEPVHVHGRKT 656  
 QY 961 KRNTNRAPQDVKFPGGQIIVG-----RRGP-----PIPKARRPESGRTWAQPGY 1003  
 |||-----RRGP-----PIPKARRPESGRTWAQPGY 716  
 Db 1004 PWPYGNKDRRSTGKSGKPGYPWPRTKRNTRRRPQDVKFPGGQIIVG-----RRGP- 1056  
 |||-----RRGP-----PIPKARRPESGRTWAQPGY 716  
 QY 717 PWPYGNKDRRSTGKSGKPGYPWPRTKRNTRRRPQDVKFPGGQIIVG-----RRGP- 776  
 |||-----RRGP-----PIPKARRPESGRTWAQPGY 716  
 Db 1057 -----PIPKARRPESGRTWAQPGYPWPYGNKDRRSTGKSGKPGYPWP 1099  
 |||-----PIPKARRPESGRTWAQPGYPWPYGNKDRRSTGKSGKPGYPWP 829  
 QY 777 LGVLATRTSIPKARRPESGRTWAQPGYPWPYGNKDRRSTGKSGKPGYPWP 829  
 |||-----PIPKARRPESGRTWAQPGYPWPYGNKDRRSTGKSGKPGYPWP 829  
 Db  
 RESULT 7  
 AAW34481  
 ID AAW34481 standard; protein; 1021 AA.  
 XX  
 AC AAW34481;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 16-MAR-1998 (first entry)  
 XX  
 DE HCV antigen combination pSOD/c200/core.  
 XX  
 KW PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;  
 KW C domain; S domain; NS5; HCV polypeptide; anti-HCV antibody; detection;  
 KW NS4.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1..902 /note= "linker"  
 FT Misc-difference 1..154 /note= "hSOD fragment"  
 FT Misc-difference 155..159 /note= "linker"  
 FT Misc-difference 160..899 /note= "c200 (amino acids 1192-1931 of HCV polypeptide)"  
 FT Misc-difference 903..1021 /note= "c22 (amino acids 2-120 of HCV polypeptide)"  
 XX



PN US5683864-A.  
XX 04-NOV-1997.  
XX 07-JUL-1992; 92US-00910760.  
XX 18-NOV-1987; 87US-00122714.  
PR 30-DEC-1987; 87US-00139886.  
PR 26-FEB-1988; 88US-00161072.  
PR 06-MAY-1988; 88US-00191263.  
PR 26-OCT-1988; 88US-00263584.  
PR 14-NOV-1988; 88US-00271450.  
PR 17-MAR-1989; 89US-00325338.  
PR 20-APR-1989; 89US-00341334.  
PR 21-APR-1989; 89US-00353896.  
PR 18-MAY-1989; 89US-00355002.  
PR 04-APR-1990; 90US-00504352.  
XX (CHIR ) CHIRON CORP.  
PA  
XX Kuo G, Houghton M, Choo Q;  
XX WPI; 1997-548976/50.  
DR N-PSDB; AAT99982.  
XX  
XX Combination of three hepatitis C virus antigens - used for detection of  
PT specific antibodies to diagnose infection.  
XX  
XX Example 6; Col 59-68; 57pp; English.  
XX  
XX This sequence represents a Hepatitis c virus (HCV) antigen combination of  
CC the invention. The HCV antigen combination comprises an antigen (Agl)  
CC comprising the C domain (i.e. amino acids (aa) 1-120 of the HCV  
CC polyprotein), or its immunologically reactive fragment containing at  
CC least 8 aa. It also comprises two additional antigens from two different  
CC polypeptide domains, including at least 8 aa from the NS3, NS4, S or NS5  
CC domains of the polyprotein, corresponding, respectively, to aa 1050-1640;  
CC 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein. Alternatively,  
CC Agl contains at least 8 aa from the 1-122 or 9-177 aa regions of the HCV  
CC polyprotein. These antigen combinations are used diagnostically to detect  
CC anti-HCV antibodies, using any standard immunoassay format. These antigen  
CC combinations have a broader range of reactivity with antibodies than any  
CC antigen individually. (Updated on 25-MAR-2003 to correct PR field.)  
XX  
XX Sequence 1021 AA;  
SQ  
Query Match 64.8%; Score 3829.5; DB 2; Length 1021;  
Best Local Similarity 67.2%; Pred. No. 5.8e-261;  
Matches 784; Conservative 24; Mismatches 79; Indels 279; Gaps 17;  
QY 1 MATKAVCVLKGDPVOGIINFEOKESGPKVWGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
DB 1 MATKAVCVLKGDPVOGIINFEOKESGPKVWGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
QY 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSTEDSVISLSDGHCIIIGRTLTV 120  
DB 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSTEDSVISLSDGHCIIIGRTLTV 120  
QY 121 HEKADLKGCGNEESTKTGNAGSLACGVIGIAQLNLSGNCNCISYPGHITGHRMAWKLS 180  
DB 121 HEKADLKGCGNEESTKTGNAGSLACGVIGIAQLNLSGNCNCISYPGHITGHRMAWKLS 180  
QY 181 AARTTSFVSLFAPGAKQNETHTVGGAAARTTSLTSLFSPGASQNLQIVDFIPVENLE 240  
DB 161 -----VDFFIPVENLE 170  
QY 241 TTRSPVFTDSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATL 300  
DB 171 TTRSPVFTDSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATL 230  
QY 301 GFQAYMSKAHGIDPNIRHTGRTTITGSPITYSTYTGKFLADGGCGGGAYDIIICDECHSTD 360  
DB 231 GFQAYMSKAHGIDPNIRHTGRTTITGSPITYSTYTGKFLADGGCGGGAYDIIICDECHSTD 290

QY 361 ATSILGIGTVLDOAETAGARLVVLTATPPGSGVTVPHPNIEEVALSTTGEIPYGAIPL 420  
DB 291 ATSILGIGTVLDOAETAGARLVVLTATPPGSGVTVPHPNIEEVALSTTGEIPYGAIPL 350  
QY 421 EVIKGGRHLIFCHSKKKCDLAAKVALGINAVAYYRGLDVSVIPSTGDVVVATDALMT 480  
DB 351 EVIKGGRHLIFCHSKKKCDLAAKVALGINAVAYYRGLDVSVIPSTGDVVVATDALMT 410  
QY 481 GYTGFDSVIDCNTCTQTVDFSLDPTFTIETITLPODAVSRTQRRGRTGRGPIYRFV 540  
DB 411 GYTGFDSVIDCNTCTQTVDFSLDPTFTIETITLPODAVSRTQRRGRTGRGPIYRFV 470  
QY 541 APCRSPGMFDSVCLCECDAGCAWYELTPAETTVRLRAYMNTPLGVPCODHLEFWEHGVF 600  
DB 471 APCRSPGMFDSVCLCECDAGCAWYELTPAETTVRLRAYMNTPLGVPCODHLEFWEHGVF 530  
QY 601 TGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSPQMDQMKCLIRLKLPTLHGPT 660  
DB 531 TGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSPQMDQMKCLIRLKLPTLHGPT 590  
QY 661 PLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS----- 696  
DB 591 PLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS----- 650  
QY 697 ---ACSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRG 748  
DB 651 VGRVLSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRG 710  
QY 749 GKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRG 803  
DB 711 AE-VIAPAVQTNQKLETFWAKHMMNFISGQYLAGLSTLPG--NPAIASLMAFTAATVS 767  
QY 804 P---DKEILYE-----AFDEMECASKAALIEEQORMAEMLKSKIQGL 843  
DB 768 PLTTSQTLFLNLILGWNVAQAAPGAATAPVAGLAGAAGTIGVGLGKVLIDILAGYGAGV 827  
QY 844 LG-----ILRRHVGPGEGAVQVMNRL 864  
DB 828 AGALVAFKMSGEVSTEDLVNLLPALSPGALVVGVCALILRRHVGPGEGAVQVMNRL 887  
QY 865 IAFASRGNHVSPTHYVPSRRLRFAQALPVNARDYNPLVETWKKPDYEPVHVHGRSSRR 924  
DB 888 IAFASRGNHVS-----GNSST- 904  
QY 925 FAQALPVNARDYNPLVETWKKPDYEPVHVHGRKTKBNTNRPPQDVKPGCGQIVG--- 981  
DB 905 -----NP-----KPK-----KKKRNTRRRPQDVKPPGGQIVGGVY 936  
QY 982 ---RRGP-----PIPKARRPEGRRTWAGQYWPWPLYGNK----- 1011  
DB 937 LLPRGPRGLVTRATKTSERSQPRRRQPIPKARPEGRRTWAGQYWPWPLYGNCGWAG 996  
QY 1012 ---DRRSTGKSGKPGKPYWPRKTKRN 1034  
DB 997 WLLSPRGRSPSGMPTD---PRRSRN 1019  
RESULT 8  
AAW40039  
ID AAW40039 standard; protein; 1021 AA.  
XX  
XX AAW40039;  
XX  
XX 26-MAY-1998 (first entry)  
XX  
XX Fusion protein c200/c22.  
XX  
XX Hepatitis C virus C domain; HCV; immunological activity; c200/c22;  
XX NS3 domain; NS4 domain; S domain; NS5 domain; fusion protein.  
XX  
XX Synthetic.  
OS  
OS Hepatitis virus.

XX US5712087-A.  
 XX 27-JAN-1998.  
 XX 12-MAY-1995; 95US-00440519.  
 XX 04-APR-1990; 90US-00504352.  
 XX 07-JUL-1992; 92US-00910760.  
 XX (CHIR ) CHIRON CORP.  
 XX Kuo G, Houghton M, Choo Q;  
 XX WPI; 1998-1119973/11.  
 XX N-PSDB; AAV09990.  
 XX Immunoassays for hepatitis C virus antibodies - using combinations of  
 XX antigenic fragments of HCV polypeptide.  
 XX Example 6; Fig 4; 59pp; English.  
 XX This sequence represents a fusion protein constructed from the hepatitis  
 XX C virus core domain (which is situated at the carboxy terminus of the  
 XX fusion protein) and a c200 construct (a fusion of the NS3 and NS3  
 XX domains). This protein used in the construction of novel combinations of  
 XX HCV antigens that have a broader range of immunological activity than any  
 XX single HCV antigen. An example of such an antigen given in this  
 XX specification comprises a first antigen containing at least 8 amino acids  
 XX of the C domain of the HCV polypeptide and a second antigen comprising at  
 XX least 8 amino acids of the NS3 domain, the NS4 domain, the S domain or  
 XX the NS5 domain of the HCV polypeptide in the form of a fusion protein, a  
 XX physical mixture or bound to a solid matrix  
 XX Sequence 1021 AA;  
 XX  
 XX Query Match 64.8%; Score 3829.5; DB 2; Length 1021;  
 XX Best Local Similarity 67.2%; Pred. No. 5.8e-261;  
 XX Matches 784; Conservative 24; Mismatches 79; Indels 279; Gaps 17;  
 XX  
 XX 1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLGHFVHFGDNTAGTS 60  
 XX 1 MATKAVCVLKGDPVQGIINFEQESNGPVKVGSIKGLTEGLGHFVHFGDNTAGTS 60  
 XX 61 AGPHFNLSRKHGPKDEERHVGDLGNVADKGVADSVIEDSVISLGDHCHIGRTLV 120  
 XX 61 AGPHFNLSRKHGPKDEERHVGDLGNVADKGVADSVIEDSVISLGDHCHIGRTLV 120  
 XX 121 HEKADDLGKGGNEESTKTGNAGSLACGVIGIAQNLSGNCISYPGHITGHRMAWKLS 180  
 XX 121 HEKADDLGKGGNEESTKTGNAGSLACGVIGIAQNLSGNCISYPGHITGHRMAWKLS 180  
 XX 181 AARTTSFVSLFAGCAQKQETHVTGGAAARTTSLTSLFSPGASQNLIVDFIPVENLE 240  
 XX 161 -----VDFFIPVENLE 170  
 XX 241 TTMRSPVFTDNSPPVVPQSFQVAHLHAPTSGKSTKVPAAAYAAQGVKVLNPSVAATL 300  
 XX 171 TTMRSPVFTDNSPPVVPQSFQVAHLHAPTSGKSTKVPAAAYAAQGVKVLNPSVAATL 230  
 XX 301 GFAYMSKAHGIDPNIRTVRTITGSPITVSTYTGKFLADGCGSGGAYDIIICDECHSTD 360  
 XX 231 GFAYMSKAHGIDPNIRTVRTITGSPITVSTYTGKFLADGCGSGGAYDIIICDECHSTD 290  
 XX 361 ATSLIGITVLDOETAGARLVLATATPPGCVTPVPHNIEVALSTTGEIPFYGKAIP 420  
 XX 291 ATSLIGITVLDOETAGARLVLATATPPGCVTPVPHNIEVALSTTGEIPFYGKAIP 350  
 XX 421 EVIKGGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVAATDALMT 480  
 XX 351 EVIKGGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVAATDALMT 410  
 XX 481 GYTGFDSVIDCNTCVTQVDFSLDPTFTIETITLPQDAVSRTQRRGRTGRGKPIYRFV 540

Db	411	GYTGFDSVIDCNTCVTQVDFSLDPTFTIETITLPQDAVSRTQRRGRTGRGKPIYRFV	470
Qy	541	APGERPSGMPDSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTPLPVCODHLEFEGV	600
Db	471	APGERPSGMPDSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTPLPVCODHLEFEGV	530
Qy	601	TGLTHIDAHFLSQTQSGENLPYLVAQVATCARAQAPPSQDWKWLRLKPTLHGPT	660
Db	531	TGLTHIDAHFLSQTQSGENLPYLVAQVATCARAQAPPSQDWKWLRLKPTLHGPT	590
Qy	661	PLLYRLGAVONEITLTHPTVKYIMTMSADLEVVT	696
Db	591	PLLYRLGAVONEITLTHPTVKYIMTMSADLEVVT	650
Qy	697	----ACSGKPAIIPDREVLYREFDEMECECSQHLPIYEQGMWLABQFKQKALGL	748
Db	651	VGRWLSGKPAIIPDREVLYREFDEMECECSQHLPIYEQGMWLABQFKQKALGL	710
Qy	749	GKPAIIPDKEVLYQOYD-----EMECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVT	803
Db	711	AE-VIAPAVQTNQKLETFWAKHMMNFIISQIYLAGLSTLPG--NPAIASLMAFTAATS	767
Qy	804	P---DKELIYE-----AFDEMECECASKAALIEBEGORMAEMLSKIQGL	843
Db	768	PLTTSQTLFNLILGGWVAQAAPGAATFVAGLAGAAGATGSLGKVLIDILAGYAGV	827
Qy	844	LG-----ILRRHVGPGEAGVOMNRL	864
Db	828	AGALVAFKINSGEVPSFEDLVNLLPALSPALVGVVCAAILRRHVGPGEAGVOMNRL	887
Qy	865	IAFASRGNHVSPTHYVPSRRAQALPVWARPDPNPLVETWKKPDYEPVHVGRSRR	924
Db	888	IAFASRGNHVSPTHYVPSRRAQALPVWARPDPNPLVETWKKPDYEPVHVGRSRR	904
Qy	925	FAQALPVWARPDPNPLVETWKKPDYEPVHVGRKTKRNTNRRPQDVKPGGGQIVG---	981
Db	905	-----NP-----KPGQ-----KKNRNTNRRPQDVKPGGGQIVG	936
Qy	982	---RRGP-----PIPKARRPEGRTWAOPGYPMPLYGNK-----	1011
Db	937	LLPRGPRLGVRATRKTSERSQPRGRQPIPKARRPEGRTWAOPGYPMPLYGNK	996
Qy	1012	---DRRSTGKSWKPGYPWPKTKRN 1034	
Db	997	WLLSPRGSRPSWGPTD---PRRSRN 1019	
RESULT 9			
ID	AAE22050	standard; protein; 1021 AA.	
XX	AAE22050;		
XX	16-JUL-2002 (first entry)		
XX	PSOD/c200/core expression plasmid protein.		
DE	Hepatitis C virus; HCV; antigen; C domain; polyprotein; NS3 domain;		
KW	NS4 domain; S domain; NS5 domain; PSOD/c200/core plasmid.		
XX	Hepatitis C virus.		
OS	Unidentified.		
OS	Chimeric.		
XX	Key	Location/Qualifiers	
FT	Region	1..154	
FT	Region	/note= "hsOD"	
FT	Region	155..159	
FT	Region	/note= "Linker region"	
FT	Region	160..899	
FT	Region	/note= "HCV c200"	
FT	Region	900..902	

FT /note= "Linker region"  
FT 903..1021  
FT /note= "HCV c22"  
XX  
PN US6312889-B1.  
PD 06-NOV-2001.  
XX  
XX 12-MAY-1995; 95US-00440549.  
XX  
XX 04-APR-1990; 90US-00504352.  
XX 07-JUL-1992; 92US-00910760.  
XX (CHIR ) CHIRON CORP.  
XX  
XX  
XX Houghton M, Choo Q, Kuo G;  
XX WPI; 2002-040268/05.  
XX N-PSDB; AAD35044.  
XX  
XX Combination of hepatitis C viral (HCV) antigens, useful in improved  
XX immunoassay for detecting HCV antibodies.  
XX  
XX Example 6; Fig 4; 58pp; English.  
XX  
XX The invention relates to combination of hepatitis C viral (HCV) antigens  
XX that have a broader range of immunological reactivity than any single HCV  
XX antigen. The combinations consist of an antigen from the C domain of the  
XX HCV polyprotein, and at least one additional HCV antigen from either the  
XX NS3 domain, the NS4 domain, the S domain, or the NS5 domain and are in  
XX the form of fusion protein, a simple physical mixture, or the individual  
XX antigens commonly bound to a solid matrix. The combinations of antigens  
XX provides broad range immunoassays for anti-HCV antibodies. The invention  
XX therefore provides a method for detecting antibodies to HCV in a mammal  
XX suspected of containing such antibodies. The present sequence is a  
XX protein encoded by pSOD/c200/core expression plasmid DNA containing HCV  
XX coding sequence  
XX  
XX Sequence 1021 AA;  
XX  
XX Query Match 64.8%; Score 3829.5; DB 5; Length 1021;  
XX Best Local Similarity 67.2%; Pred. No. 5.8e-261;  
XX Matches 784; Conservative 24; Mismatches 79; Indels 279; Gaps 17;  
XX  
XX 1 MATKAVCVLKGDCPVQGIINFEOKESGPKVKGSIKGLTEGLHGHVHFEGDNTAGCTS 60  
XX 1 MATKAVCVLKGDCPVQGIINFEOKESGPKVKGSIKGLTEGLHGHVHFEGDNTAGCTS 60  
XX  
XX 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIGRTLTV 120  
XX 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIGRTLTV 120  
XX  
XX 121 HEKADLKGKGNBESTKTGNAGSRLACGVGTGIAQNLNSGCNCSYIPGHITGHRMAWKLS 180  
XX 121 HEKADLKGKGNBESTKTGNAGSRLACGVGTGIAQNLNSGCNCSYIPGHITGHRMAWKLS 180  
XX  
XX 181 AARTSGFVSLFAPGAKQNETHTVGGAAARTSLTSLFSPGASQNIQLIVDFIPVENLE 240  
XX 161 -----VDFIPVENLE 170  
XX  
XX 241 TTMRSPTVFTDNSPPVVPQSFQVAHLHAPTSGKSTKVPAAAYAAQGYKVLVLPNSVAATL 300  
XX 171 TTMRSPTVFTDNSPPVVPQSFQVAHLHAPTSGKSTKVPAAAYAAQGYKVLVLPNSVAATL 230  
XX  
XX 301 GFAYMSKAHGDIPNIRTKRTITGSPITYSTYVGKFLADGGCGSGAYDIIICDECHSTD 360  
XX 231 GFAYMSKAHGDIPNIRTKRTITGSPITYSTYVGKFLADGGCGSGAYDIIICDECHSTD 290  
XX  
XX 361 ATSLIGTGLDQAEATAGARLVLATATPGSVTPHPNIEEVALSTTGEIPFGKAIP 420  
XX 291 ATSLIGTGLDQAEATAGARLVLATATPGSVTPHPNIEEVALSTTGEIPFGKAIP 350  
XX  
XX 421 EVIKGGRHLIFCHSKKKCDLAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 480

351 EVIKGGRHLIFCHSKKKCDLAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 410  
481 GYTGPDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTQRRGTRGKPGIYRFV 540  
411 GYTGPDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTQRRGTRGKPGIYRFV 470  
541 AGERPSGMFSDSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCDHLEFEGVF 600  
471 AGERPSGMFSDSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCDHLEFEGVF 530  
601 TGLTHIDAHFLSQTQSGENLPIYLVAQYATVCARQAAPPSPDQMWKCLIRLKPTHGPT 660  
531 TGLTHIDAHFLSQTQSGENLPIYLVAQYATVCARQAAPPSPDQMWKCLIRLKPTHGPT 590  
661 PLYLRGAVONEITLTHPVTKYIMTMSADLEVTVS----- 696  
591 PLYLRGAVONEITLTHPVTKYIMTMSADLEVTVS----- 650  
697 ----ACSGKPAIIPDREVLYREFDEMEECSSQHLPIYIEQGMMLAEQFKQKALGL-----SRG 748  
651 VGRVVLGKPAIIPDREVLYREFDEMEECSSQHLPIYIEQGMMLAEQFKQKALGLQTSRQ 710  
749 GKPALVPDKVLYQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVVT 803  
711 AE-VIAPAVQTNQKLETFWAKHMNFISGIQVLAGLSTLPG--NPAAIASLMAFTAATVS 767  
804 P----DKEIILYE-----AFDEMEECASKAALIEEGORMAEMLKSKIQGL 843  
768 PLTTSQTLILNIGGHWAAQLAAPGAATAFVGAGLAGAIGSVGLGKVLIDILLAGYGAGV 827  
844 LG-----ILRRHVGPGEAGVQVMNRL 864  
828 AGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQVMNRL 887  
865 IAFASGNHVSPTHYVPSRRRPAQALPVWARPDPYNPPLVETWKKPDYEPVHVHSSRR 924  
888 IAFASGNHVSP-----GNSST- 904  
925 FAQALPVWARPDPYNPPLVETWKKPDYEPVHVHGRKTKRNTNRRPQDVKPPGGQIVG--- 981  
905 -----NP-----KPO-----KQKRNTRNRRPQDVKPPGGQIVGGVY 936  
982 ----RRGP-----PIPKARPEGRTPAQGYPWPLYGNK----- 1011  
937 LLPRGPRGLGVRATRKTSERSQPRRRQPIPKARPEGRTPAQGYPWPLYGNCGCWAG 996  
1012 ----DRRSTGSKGKPGYWPWPKTKRN 1034  
997 WLLSPRGRSPSMGPTD---PRRRSRN 1019  
RESULT 10  
AAR68547  
ID AAR68547 standard; protein; 841 AA.  
XX  
XX AAR68547;  
XX AC  
XX 25-MAR-2003 (revised)  
XX 17-AUG-1995 (first entry)  
XX DT  
XX HCV protease/hSOD fusion protein expression vector cflSODp600.  
XX DE  
XX KW Hepatitis C virus protease/hSOD fusion protein; HCV;  
XX expression vector cflSODp600; viral infectivity inhibition.  
XX OS  
XX Hepatitis C virus.  
XX PH Key Location/Qualifiers  
XX FT Peptide 1..155  
XX /label= SOD leader  
XX PN US5371017-A.



SQ Sequence 841 AA;

Query Match 51.6%; Score 3050.5; DB 6; Length 841;  
Best Local Similarity 72.6%; Pred. No. 4.2e-206;  
Matches 615; Conservative 10; Mismatches 37; Indels 185; Gaps 11;

Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
Db 1 MATNPVCLVKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60

Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIGRTLVV 120  
Db 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIGRTLVV 120

Qy 121 HEKADDLGKGNBESTKTGNAGSLACGVIGIAQLNSGCNCSTYPCGHITGHR----- 173  
Db 121 HEKADDLGKGNBESTKTGNAGSLACGVIGIR-----IGTVYV-NHLTPLRDWAHNGL 174

Qy 174 -----NAWKLGSA-----RTTSGFVS----- 190  
Db 175 RDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPADGWVSKGWR 234

Qy 191 LFAP-----GAKQNETH-----VTG 205  
Db 235 LLAPITAYAAQOTRGLLGCIIITSLTRDKNQVEGEVQIVSTAQTFLATCIINGVCMTYYH 294

Qy 206 GAAART-----TSGLT----- 216  
Db 295 GAGTRTITASPKGPVIQMYTNDQDLVGPASQGRSLTPTCGSSDDLYLVTRIADVIPVR 354

Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241  
Db 355 RRGDSRGLLSPRFISYLGSSGGPLCPAGHVGIFRAAVCTRGVAKAVDFIPVENLET 414

Qy 242 TMRSPVTDNSSPPVPSQVQVAHLHAPTSQSGKSTKVPAAQGYKVLVLPNSVAATLG 301  
Db 415 TMRSPVTDNSSPPVPSQVQVAHLHAPTSQSGKSTKVPAAQGYKVLVLPNSVAATLG 474

Qy 302 FGAYMSKAGIDPNIRGTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 361  
Db 475 FGAYMSKAGIDPNIRGTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 534

Qy 362 TSILGIGTVLDQAEAGARLVWLATATPPGSVTVVPHENIEEVALSTTGEIPFYGKAIPLE 421  
Db 535 TSILGIGTVLDQAEAGARLVWLATATPPGSVTVVPHENIEEVALSTTGEIPFYGKAIPLE 594

Qy 422 VIKGGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 481  
Db 595 VIKGGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 654

Qy 482 YTGDFDSVIDNCVTQTVDFSLDPTFTIITLPODAVSRTPRGRTGRKPGIYRFVA 541  
Db 655 YTGDFDSVIDNCVTQTVDFSLDPTFTIITLPODAVSRTPRGRTGRKPGIYRFVA 714

Qy 542 PGRSPGMFDSVLCYDAGCAWYELTPAETTVRLRAYWNTTGLPVCQDHLFEWGVFT 601  
Db 715 PGRSPGMFDSVLCYDAGCAWYELTPAETTVRLRAYWNTTGLPVCQDHLFEWGVFT 774

Qy 602 GLTHIDAHFISQTKSGENLPYLVAQATVCARAQAPPPSDQWQKCLIRLKPTLHGPTP 661  
Db 775 GLTHIDAHFISQTKSGENLPYLVAQATVCARAQAPPPSDQWQKCLIRLKPTLHGPTP 834

Qy 662 LLYRLGA 668  
Db 835 LLYRLGA 841

RESULT 12

ID ADA07875 standard; protein; 841 AA.

XX

AC

XX

DT 06-NOV-2003 (first entry)

XX HCV protease/hsOD fusion protein.

XX HCV; virucide; NS3 protease; serine protease; hsOD; superoxide dismutase;  
KM yeast a-factor; interleukin-2S; ubiquitin; beta-galactosidase;  
KW beta-lactamase; horseradish peroxidase; glucose oxidase; urease;  
KW HCV infection; cfISODp600; human.

XX Chimeric.

OS Hepatitis C virus.

OS Homo sapiens.

XX US2003064499-A1.

PN 03-APR-2003.

PD 18-JUN-2001; 2001US-00884455.

XX 04-APR-1990; 90US-00505433.

PR 04-APR-1991; 91US-00680296.

PR 06-DEC-1994; 94US-00350884.

PR 12-MAY-1995; 95US-00440548.

PR 06-SEP-1996; 96US-00709177.

PR 18-FEB-1999; 99US-00253675.

XX (HOUG/) HOUGHTON M.  
PA (CHOO/) CHOO Q.  
PA (KUOG/) KUO G.

XX Houghton M, Choo Q, Kuo G;  
PI WPI; 2003-540789/51.  
DR N-PSDB; ADA07874.

XX A composition for assaying and designing antiviral agents specific for  
PT Hepatitis C virus (HCV) comprises a purified proteolytic polypeptide from  
PT HCV or a polynucleotide which encodes HCV protease.  
PS Disclosure; Fig 10; 40pp; English.

XX The invention relates to a composition comprising a purified proteolytic  
CC polypeptide derived from Hepatitis C virus (HCV) or a polynucleotide  
CC which encodes only the HCV protease or an active HCV protease analogue,  
CC or which encodes a fusion protein comprising HCV protease or HCV protease  
CC analogue, and a fusion partner. Also included are a fusion protein  
CC comprising a fusion partner fused to a proteolytic polypeptide derived  
CC from HCV, a method for assaying compounds for activity against HCV  
CC with a compound capable of inhibiting serine protease activity and  
CC measuring inhibition of the proteolytic activity of the HCV protease) and  
CC an expression vector for producing HCV protease or HCV protease analogues  
CC in a host cell (comprising a polynucleotide encoding HCV protease or an  
CC HCV protease analogue, transcriptional and translational regulatory  
CC sequences functional in the host cell operably linked to the HCV protease  
CC encoding polynucleotide and a selectable marker). The fusion partner is  
CC selected from hsOD (human superoxide dismutase), yeast a-factor,  
CC interleukin (IL)-2S, ubiquitin, beta-galactosidase, beta-lactamase,  
CC horseradish peroxidase, glucose oxidase and urease. The composition is  
CC useful in assaying and designing antiviral agents specific for HCV. The  
CC method is used in identifying antiviral agents effective for treating  
CC HCV. The present sequence is an HCV protease/hsOD fusion protein.

XX Sequence 841 AA;

Query Match 51.6%; Score 3050.5; DB 7; Length 841;

Best Local Similarity 72.6%; Pred. No. 4.2e-206;

Matches 615; Conservative 10; Mismatches 37; Indels 185; Gaps 11;

Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60

Db 1 MATNPVCLVKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60

```
Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKQGVADVSIEDSVISLSDHCHIIIGRTLIV 120
Db 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKQGVADVSIEDSVISLSDHCHIIIGRTLIV 120
Qy 121 HEKADDLGKGNNESTKTGNAGSLACGVIGIAQNLNSGCNCIYPGHITGHR----- 173
Db 121 HEKADDLGKGNNESTKTGNAGSLACGVIGIR-----IGTYVI-NHLTPLRDWAHNGL 174
Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190
Db 175 RDLAVAVEPVFSQMETKLIITWADTAACDGIINGLPVSARRREILLGPADGNVSKGR 234
Qy 191 LFAP-----GAKONETH-----VTG 205
Db 235 LLAPITAYAOQTRGLLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCIINGVCWTYH 294
Qy 206 GAAART-----TSGLT----- 216
Db 295 GAGRTTASPKGPVIQMTNVDDLVGWPAQSGTRSLTPCTCGSSDLYLVRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Db 355 RRGDSRGLSPRISYLVKGSGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPVFTDNSSPPVQSFQVAHLHAPTGSCKTKVPAAYAAQYKVLNPSVAATLG 301
Db 415 TMRSPVFTDNSSPPVQSFQVAHLHAPTGSCKTKVPAAYAAQYKVLNPSVAATLG 474
Qy 302 FGAYMSKAHGDPIRITCVRTITTSPTITYSTYCKFLADGCGSGAYDIIICDCHSTDA 361
Db 475 FGAYMSKAHGDPIRITGVRTITTSPTITYSTYCKFLADGCGSGAYDIIICDCHSTDA 534
Qy 362 TSIILGIGTVLDOAETAGARLVLTATPPGVTVPHPNIEEVALSTTGEIPFYKAIPL 421
Db 535 TSIILGIGTVLDOAETAGARLVLTATPPGVTVPHPNIEEVALSTTGEIPFYKAIPL 594
Qy 422 VIKGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVATDALMTG 481
Db 595 VIKGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVATDALMTG 654
Qy 482 YTGFDSDVIDCNTCTQTVDFSLDPTFTIETILPQDAVSRTORGRGKPGIYRVA 541
Db 655 YTGFDSDVIDCNTCTQTVDFSLDPTFTIETILPQDAVSRTORGRGKPGIYRVA 714
Qy 542 PGERPSGMFDSVLCECYDAGCAYELTPAETTVRLRAYMTTGLPVQCQDHLFEWEGVFT 601
Db 715 PGERPPGMFDSVLCECYDAGCAYELTPAETTVRLRAYMTTGLPVQCQDHLFEWEGVFT 774
Qy 602 GLTHIDHFLSQTQSGENLPYLVAQATVCARQAQPPPSWDQMKLIRLKLPTLHGPTP 661
Db 775 GLTHIDHFLSQTQSGENLPYLVAQATVCARQAQPPPSWDQMKLIRLKLPTLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841
```

## RESULT 13

AAW01701

ID AAW01701 standard; protein; 841 AA.

XX

AC

AAW01701;

XX

DT 17-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 03-APR-1997 (first entry)

XX

DE hSOD-HCV fusion protein.

XX

KW HCV; NS3; non-structural domain 3; protease; polyprotein; inhibitor;

XX screen; processing; infection; treatment; probe; hepatitis C virus.

OS

Hepatitis C virus; Virus.

OS Homo sapiens.  
XX Chimeric.FH Key Location/Qualifiers  
FT Protein 156..841  
FT /label= HCV\_protease

XX US5585258-A.

XX 17-DEC-1996.

XX 06-DEC-1994; 94US-00350884.

XX 04-APR-1990; 90US-00505433.

XX 04-APR-1991; 91US-00680296.

XX (CHIR ) CHIRON CORP.

XX Choo Q, Kuo G, Houghton M;

XX WPI; 1997-051175/05.  
XX N-PSDB; AAT59261.XX Compn. contg. hepatitis C virus NS3 domain protease and related fusion  
XX proteins - useful for screening specific inhibitors, potential antiviral  
XX agents, prepn. of antibodies and for cleaving specific poly.peptide(s).  
XX Example 4; Col 77-84; 68pp; English.XX Compns. comprising the hepatitis C virus (HCV) NS3 domain protease or  
XX its active truncation analogues are claimed. Also new are fusion proteins  
XX comprising the protease (or analogues) and, e.g. human superoxide (SOD)  
XX or ubiquitin. The protease is essential for polypeptide processing, and  
XX thus infectivity, in HCV. The compns. are used to screen for specific  
XX inhibitors (possibly useful as antiviral agents), to generate specific  
XX antibodies and to cleave specific polypeptides. HCV cDNA clones (AAT59250  
XX - 56 encoding AAW01686-92 resp.) were isolated from HCV genomic library  
XX using probes AAT59244-49. The clones were used in the preparation of full  
XX -length SOD-protease fusion proteins. The present sequence is encoded by  
XX vector cFISODp600 which contains a full-length HCV protease coding  
XX sequence fused to a functional hSOD leader. The resulting vector encodes  
XX amino acids 1-151 of hSOD, and amino acids 946-1630 of HCV (corresponding  
XX to 1-686 of AAW01693). (Updated on 25-MAR-2003 to correct PF field.)  
XX (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 841 AA;

Query Match 51.5%; Score 1047.5; DB 2; Length 841;

Best Local Similarity 72.6%; Pred. No. 6.8e-206;

Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;

Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60

Db 1 MATNPVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60

Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKQGVADVSIEDSVISLSDHCHIIIGRTLIV 120

Db 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKQGVADVSIEDSVISLSDHCHIIIGRTLIV 120

Qy 121 HEKADDLGKGNNESTKTGNAGSLACGVIGIAQNLNSGCNCIYPGHITGHR----- 173

Db 121 HEKADDLGKGNNESTKTGNAGSLACGVIGIR-----GTYVI-NHLTPLRDWAHNGL 174

Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190

Db 175 RDLAVAVEPVFSQMETKLIITWADTAACDGIINGLPVSARRREILLGPADGNVSKGR 234

Qy 191 LFAP-----GAKONETH-----VTG 205

Db 235 LLAPITAYAOQTRGLLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCIINGVCWTYH 294

Qy 206 GAAART-----TSGLT----- 216

Db 295 GAGTRTTIASPKGPVIQMYTNVDQLVGPASQGTSLTPCTCGSSDLYLVTRHADVIPVR 354  
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241  
Dy 355 RRGDSRGLSPRPISYLVKSGSGPGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLET 414  
Qy 242 TMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLG 301  
Dy 415 TMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLG 474  
Qy 302 FGAYMSKAHGIDPNIRGTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDA 361  
Dy 475 FGAYMSKAHGIDPNIRGTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDA 534  
Qy 362 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 421  
Dy 535 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 594  
Qy 422 VIKGRHLIFCHSKKCKDELAALVALGINAVAYYRGLDVSVIPTSGDGVVVVATDALMTG 481  
Dy 595 VIKGRHLIFCHSKKCKDELAALVALGINAVAYYRGLDVSVIPTSGDGVVVVATDALMTG 654  
Qy 482 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOGRTGRKPGIYRFVA 541  
Dy 655 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOGRTGRKPGIYRFVA 714  
Qy 542 PGERPSPGFSSVLCCEYDAGCANYELTPAETTVRLRAYNMTPLPVCODHLEFEGVFT 601  
Dy 715 PGERPSPGFSSVLCCEYDAGCANYELTPAETTVRLRAYNMTPLPVCODHLEFEGVFT 774  
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMKLIRLKLPHGTP 661  
Dy 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMKLIRLKLPHGTP 834  
Qy 662 LLYRLGA 668  
Dy 835 LLYRLGA 841

RESULT 14  
AAW46397 standard; protein; 841 AA.  
AC AC  
XX AAW46397;  
XX  
DT 27-AUG-2003 (revised)  
DT 07-MAY-1998 (first entry)  
XX  
DE Amino acid sequence of the vector cf1SODp600.  
XX  
XX Protease; HCV; NS3 domain; human superoxide dismutase; fusion protein;  
KW assay; activity; anti-HCV.  
XX  
XX Synthetic.  
OS Hepatitis C virus.  
OS Homo sapiens.  
XX  
XX US5712145-A.  
XX  
XX 27-JAN-1998.  
XX  
XX 06-SEP-1996; 96US-00709173.  
XX  
XX 04-APR-1990; 90US-00505433.  
PR 04-APR-1991; 91US-00680296.  
PR 06-DEC-1994; 94US-00350884.  
PR 12-MAY-1995; 95US-00440548.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Choo Q, Kuo G, Houghton M;  
XX WPI; 1998-119986/11.  
DR

DR N-PSDB; AAV04993.  
XX Recombinant hepatitis C virus protease - useful in screening drugs for  
PT activity against hepatitis C virus.  
PT  
XX Disclosure; Fig 10A-G; 68pp; English.  
PS  
XX The present sequence represents the amino acid sequence of the vector  
XX cf1SODp600. This vector contains a full length Hepatitis C virus (HCV)  
CC protease coding sequence fused to a functional human superoxide dismutase  
CC leader. The vector was used to express the protease fusion protein in  
CC Escherichia coli. The HCV protease is believed to cleave itself from the  
CC genomic polyprotein. In the absence of protease activity, the HCV  
CC polyprotein should remain in its unprocessed form, and thus render the  
CC virus non-infectious. Inhibitors of protease activity should also inhibit  
CC viral infectivity. The protease can therefore be used for assaying  
CC compounds for activity against HCV. (Updated on 27-AUG-2003 to correct OS  
CC field.)  
XX  
SQ Sequence 841 AA;  
Query Match 51.5%; Score 3047.5; DB 2; Length 841;  
Best Local Similarity 72.6%; Pred. No. 6.8e-206;  
Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;  
Qy 1 MATKAVCLVKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Dy 1 MATNPVCLVKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADSVIEDSVISLGDHICIGRTLTV 120  
Dy 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADSVIEDSVISLGDHICIGRTLTV 120  
Qy 121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIQAQNLNSGCNCSYIPGHITGHR 173  
Dy 121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIR-----GTIVY-NHLTPLRDWAHNG 174  
Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190  
Dy 175 RDLAVAVEPVVFSOMETKLITWGAADTAACGDIINGLIPVSARRREILLGPADGMVSKGWR 234  
Qy 191 LFAP-----GAKQNEH-----VTG 205  
Dy 235 LLAPITAYAQOTRGLLCIITSLTRDKQVEGEVQIVSTAQTFLATCIINGVCWTYH 294  
Qy 206 GAAART-----TSGLT----- 216  
Dy 295 GAGTRTIASPKGPVIQMYTNVDQLVGPASQGTSLTPCTCGSSDLYLVTRHADVIPVR 354  
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241  
Dy 355 RRGDSRGLSPRPISYLVKSGSGPGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLET 414  
Qy 242 TMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLG 301  
Dy 415 TMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLG 474  
Qy 302 FGAYMSKAHGIDPNIRGTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDA 361  
Dy 475 FGAYMSKAHGIDPNIRGTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDA 534  
Qy 362 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 421  
Dy 535 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 594  
Qy 422 VIKGRHLIFCHSKKCKDELAALVALGINAVAYYRGLDVSVIPTSGDGVVVVATDALMTG 481  
Dy 595 VIKGRHLIFCHSKKCKDELAALVALGINAVAYYRGLDVSVIPTSGDGVVVVATDALMTG 654  
Qy 482 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOGRTGRKPGIYRFVA 541  
Dy 655 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOGRTGRKPGIYRFVA 714



```
Qy 542 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFEGVFT 601
Db 715 PGERPPGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFEGVFT 774
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPFLHGPTP 661
Db 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPFLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841

RESULT 15
AAW97609
ID AAW97609 standard; protein; 841 AA.
XX
AC AAW97609;
XX
DT 26-MAY-1999 (first entry)
XX
DE Amino acid sequence of vector cf15ODp600.
XX
KW HCV NS3 protease; truncation analog; HCV control; protease activity;
KM viral infectivity; inactive non-cleaving protease.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN USS885799-A.
XX
PD 23-MAR-1999.
XX
PF 06-SEP-1996; 96US-00709177.
XX
PR 04-APR-1990; 90US-00505433.
PR 04-APR-1991; 91US-00680296.
PR 06-DEC-1994; 94US-00350884.
PR 12-MAY-1995; 95US-00440548.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Choo Q, Kuo G, Houghton M;
XX
PT WPI; 1999-228536/19.
XX
DR N-PSDB; AAX26398.
XX
PT Preparation of new Hepatitis C Virus NS3 protease - useful for screening
XX for compounds which inhibit HCV infectivity.
XX
PS Example 3; Fig 10; 71pp; English.
XX
CC The specification describes a method for making a purified Hepatitis C
CC virus (HCV) NS3 protease or active truncation analog. If the HCV protease
CC N-terminal cleavage signal is excluded (so that self-cleavage is
CC prevented), the HCV protease remains in its unprocessed form, and renders
CC the virus noninfectious. The protease is therefore useful for assaying
CC pharmaceutical agents for control of HCV, as compounds which inhibit
CC protease activity sufficiently will also inhibit viral infectivity. An
CC inactive non-cleaving protease can be used to screen for inhibitors.
CC Recombinant expression systems can be utilised to prepare recombinant HCV
CC which can be used to produce monoclonal antibodies. The present sequence
XX was created in the course of the invention
XX
SQ Sequence 841 AA;

Query Match 51.5%; Score 3047.5; DB 2; Length 841;
Best Local Similarity 72.6%; Pred. No. 6.8e-206;
Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;

Qy 1 MATKAVCVLKGDPVQGIINFRQKESNGPVKVGWSIKGLTEGLHGFHVHFGDNTAGCTS 60
Db 1 MATNPVCLVKGDPVQGIINFRQKESNGPVKVGWSIKGLTEGLHGFHVHFGDNTAGCTS 60
```

```
Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCLIGRTLAV 120
Db 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCLIGRTLAV 120
Qy 121 HEKADDLGKGNBEESTKTGNAGSRLACGVIGIAGNLNCGNCSTIYPGHITGHR 173
Db 121 HEKADDLGKGNBEESTKTGNAGSRLACGVIGIR-----GTYYV-NHLPLRDMAHNGL 174
Qy 174 -----MAWKLGSA-----RTSGFVS----- 190
Db 175 RDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLFPVSARRGREILLGPADGMVSKGWR 234
Qy 191 LFAP-----GAKONETH-----VTG 205
Db 235 LLAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCIINGVCWTYYH 294
Qy 206 GAART-----TSLGT----- 216
Db 295 GAGTRTIASPKGPVIQMYTNVDQDLVGNPASPQSTRSLTPTCGSSDLYLVTRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Db 355 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPTFTDNSSPPVFPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNLPSSVAATLG 301
Db 415 TMRSPTFTDNSSPPVFPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNLPSSVAATLG 474
Qy 302 FGAYMSKAHGIDPNIRTVRTITTSPTTSTYTKFLADGGCGGAYDIIICDECHSTDA 361
Db 475 FGAYMSKAHGIDPNIRTVRTITTSPTTSTYTKFLADGGCGGAYDIIICDECHSTDA 534
Qy 362 TSLIGIGTVLDQAEATAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 421
Db 535 TSLIGIGTVLDQAEATAGARLVLTATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 594
Qy 422 VIKGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVSVITSGDVVVVATDALMTG 481
Db 595 VIKGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVSVITSGDVVVVATDALMTG 654
Qy 482 YTGDFDSVIDCNTCTQTVDFSLDPTFTTITITLPQDAVSRTQRRGTRGKPGIYRFVA 541
Db 655 YTGDFDSVIDCNTCTQTVDFSLDPTFTTITITLPQDAVSRTQRRGTRGKPGIYRFVA 714
Qy 542 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFEGVFT 601
Db 715 PGERPPGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFEGVFT 774
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPFLHGPTP 661
Db 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPFLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841
```

Search completed: November 7, 2005, 20:10:27  
Job time : 136.969 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:00:21 ; Search time 22.7031 Seconds  
(without alignments)  
4657.604 Million cell updates/sec

Title: US-10-658-782-6  
Perfect score: 5912  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRTGKSGKGPWP 1099

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2909.5	49.2	3011	1 GNWVC3	genome polyprotein
2	2850	48.2	3011	1 S40770	genome polyprotein
3	2821	47.7	3011	1 GNWVCH	genome polyprotein
4	2768.5	46.8	3010	1 A45573	genome polyprotein
5	2746	46.4	3010	1 GNWVCJ	genome polyprotein
6	2741.5	46.4	3010	1 GNWVTC	genome polyprotein
7	2737.5	46.3	3010	1 S18030	genome polyprotein
8	2725.5	46.1	3010	1 GNWVTW	genome polyprotein
9	2455.5	41.5	3014	1 JCS620	genome polyprotein
10	2431	41.1	3033	1 JQ1303	genome polyprotein
11	2428	41.1	3033	1 GNWVJ8	genome polyprotein
12	1980.5	33.5	386	2 S68016	genome polyprotein
13	1966.5	33.3	876	2 PC2219	polyprotein - hepa
14	1483.5	25.1	492	2 PS0326	polyprotein - hepa
15	1386.5	23.5	716	2 JQ1366	polyprotein - hepa
16	1067	18.0	216	2 S21337	genome polyprotein
17	1001	16.9	194	2 S06067	nonstructural prot
18	998	16.9	194	2 A54317	probable nonstruct
19	980	16.6	182	2 S32748	genome polyprotein
20	947	16.0	184	2 A61196	genome polyprotein
21	834.5	14.1	3005	2 T08841	polyprotein - dour
22	825.5	14.0	2970	2 T08839	polyprotein - marm
23	821	13.9	154	1 DSHUC2	superoxide dismuta
24	807	13.7	179	4 T43640	superoxide dismuta
25	720	12.2	135	2 PS0327	polyprotein - hepa
26	719	12.2	135	2 PS0328	polyprotein - hepa
27	710	12.0	209	2 PC1306	genome polyprotein
28	693	11.7	135	2 PS0329	polyprotein - hepa
29	684.5	11.6	152	2 S36108	superoxide dismuta

30	680	11.5	154	2 JQ0915	superoxide dismuta
31	678	11.5	154	2 JC1192	superoxide dismuta
32	667	11.3	154	1 DSHOC2	superoxide dismuta
33	659.5	11.2	152	2 S33162	superoxide dismuta
34	652	11.0	152	1 DSB0C2	superoxide dismuta
35	650.5	11.0	152	1 DSPGC2	superoxide dismuta
36	635	10.7	151	2 A24475	superoxide dismuta
37	599	10.1	153	2 S65436	superoxide dismuta
38	589.5	10.0	166	2 S29782	superoxide dismuta
39	559.5	9.5	152	2 S04623	superoxide dismuta
40	536.5	9.1	151	1 DSWFC2	superoxide dismuta
41	533	9.0	151	1 S09568	superoxide dismuta
42	521	8.8	150	2 S05021	superoxide dismuta
43	512	8.7	153	2 A45171	superoxide dismuta
44	505	8.5	153	2 S48117	superoxide dismuta
45	493	8.3	153	1 DSFFC2	superoxide dismuta

ALIGNMENTS

RESULT 1

GNWVC3

Genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.-L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: UNIPROT:P26664; GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DDBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>

F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F:1230-1237/Product: hepatitis virus #status predicted <NS3>

F:1312-1317/Region: nucleotide-binding motif A (P-loop)

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 49.2%; Score 2909.5; DB 1; Length 3011;  
Best Local Similarity 53.3%; Pred. No. 4e-173;  
Matches 622; Conservative 33; Mismatches 99; Indels 419; Gaps 16;

QY 224 SONTQLIVDFIPVENLETTMRSPPVFTDSSPPVPSQVAHLHAPTCGKTKVPAYAY 283

Db	1186	TRGAKAVDPIPVENLETTMRSPVFTDNSSPPVVFQVAHLHPTGSGSKTKVPAAYA	1245
Qy	284	AQGYKVLVNLPSVAATLGFAYMSKAHGIDPNIRTVGRTITTTGSPITYSTYGRFLADGGC	343
Db	1246	AQGYKVLVNLPSVAATLGFAYMSKAHGIDPNIRTVGRTITTTGSPITYSTYGRFLADGGC	1305
Qy	344	SGGAYDIIICDECHSDTATSILGIGTVLDQAEATAGARLVVLATATPPGVTVPHPNIEEV	403
Db	1306	SGGAYDIIICDECHSDTATSILGIGTVLDQAEATAGARLVVLATATPPGVTVPHPNIEEV	1365
Qy	404	ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV	463
Db	1366	ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV	1425
Qy	464	IPTSGDVVVVATDALMTGYTGDSDVIDCNTCTVQTVDRLDPTFTIETITLQDQAVSRT	523
Db	1426	IPTSGDVVVVATDALMTGYTGDSDVIDCNTCTVQTVDRLDPTFTIETITLQDQAVSRT	1485
Qy	524	QRRGRTGRGPGIYRFVAPGERPSGMFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNT	583
Db	1486	QRRGRTGRGPGIYRFVAPGERPSGMFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNT	1545
Qy	584	PGLPVQODHLEFEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPSWD	643
Db	1546	PGLPVQODHLEFEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPSWD	1605
Qy	644	QMKKILRLKPTLHGPTPLLYRLGAVONEITLTHPTVKYIMTQMSADLEVVTSS-----	696
Db	1606	QMKKILRLKPTLHGPTPLLYRLGAVONEITLTHPTVKYIMTQMSADLEVVTSSWVLVVG	1665
Qy	697	-----ACSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLA	735
Db	1666	VLAALAAAYCLSTGCWIVGVRVLSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLA	1725
Qy	736	EQFKQKALGL-----SRGKGPAIVPDKEVLYQQVD-----EMECSSQAAPYIEQAQVIAHQ	786
Db	1726	EQFKQKALGLQLOTSARQAE-VIAPAVQTNWQKLETFWAKHMMFISGIGYLAGLSTLPG-	1783
Qy	787	FKEKVLGLIDNQVWVTP---DKELIYE-----AFDEMECASKAALI	826
Db	1784	-NPAIASLMAFTAAVYSPLTTSQTLFLNIIIGWVAAQLAAPGAATAPVAGLAGAAGISV	1842
Qy	827	EEGQRMAELKSKIQGLLG-----V-----IL	847
Db	1843	GLCKVLIDILAGYGAGVAGALVAFKIMSGEVSPFEDLVNLLPALSPALVVGVCAAIL	1902
Qy	848	RRHVGPGEQAVQWNRLLIAPASRGNHVSPTHYPSS-----	882
Db	1903	RRHVGPGEQAVQWNRLLIAPASRGNHVSPTHYPSPESDAAAARVTAISSLVTQLLRRLHQ	1962
Qy	883	-----	882
Db	1963	WISSECTTPCGSGLRWLDIWDNI CEVLSDFTWLKAKLMPOLPGIPFVSCQGYKGVWRVD	2022
Qy	883	-----	882
Db	2023	GIMHTRCHCAEITGHVQNGTMRIVGPRTCRNWMSGTFPINAYTTGCTPLPAPNYTFAL	2082
Qy	883	-----RSRRFA-----	888
Db	2083	WRVSAEYVEIRQVDGPHYVTGTTDNLKCPQVPSPEFFTELDGVRHLHRFAPCKPLL	2142
Qy	889	-----QALPVWARPDY-----NPP-----	902
Db	2143	EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSSHITAEAGRRLARGSPSPVASS	2202
Qy	903	-----LVET---WKK-----PPEYPPVWH	918
Db	2203	ASQLSAPSLKATCTANHDSQDAELIEANLLWRQEMGNNITRVESENKVVILDSFDPLVAE	2262
Qy	919	G-----RSSRRFAQALPVWARPDYNPPLVETWKKPDYEPVPPVHGRTKRNTNR	966
Db	2263	EDBREISVPAEILKSRRFAQALPVWARPDYNPPLVETWKKPDYEPVPPVH-----	2312

Qy	967	RPQDVKPGCGQIVGRRGPPPIKARR	992
Db	2313	-----GCPLPPEPKSPVPVPPRK	2329
RESULT 2			
S40770			
genome polyprotein - hepatitis C virus			
N;Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu			
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C;Species: hepatitis C virus			
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004			
C;Accession: S40770; PC1285			
R;Okamoto, H.			
submitted to the EMBL Data Library, March 1992			
A;Reference number: S40770			
A;Accession: S40770			
A;Molecule type: genomic RNA			
A;Residues: 1-3011 <OKA>			
A;Cross-references: UNIPROT:Q03463; EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g22158			
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tauda,			
Jpn. J. Exp. Med. 60, 167-177, 1990			
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.			
A;Reference number: PC1284; MUID:91013116; PMID:2170712			
A;Accession: PC1285			
A;Molecule type: genomic RNA			
A;Residues: 1-513 <OK2>			
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512			
A;Experimental source: isolate HC-J1			
C;Superfamily: hepatitis C virus genome polyprotein			
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine			
F;2-115/Product: capsid protein C #status predicted <CPC>			
F;116-191/Product: envelope protein M #status predicted <EPM>			
F;192-389/Product: major envelope protein E #status predicted <MEE>			
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>			
F;1007-1615/Product: hepatitisvirin #status predicted <NS3>			
F;1230-1237/Region: nucleotide-binding motif A (P-loop)			
F;1312-1317/Region: nucleotide-binding motif B			
F;1316-1319/Region: DEXH motif			
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>			
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>			
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>			

Query Match.	48.2%	Score	2850	DB 1	Length	3011	
Best Local Similarity	51.2%	Pred. No.	2.1e-169	Indels	420	Gaps	18
Matches	608	Conservative	44	Mismatches	115		
Qy	204	TGGAARTTSGTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNSSPPVPPQSFO	262				
Db	1165	SGGPLLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLETTMRSPVFTDNSSPPAVPQSFO	1224				
Qy	263	VAHLHAPTSGSKTKVPAAAYAAQGYKVLVNLPSVAATLFGAYMSKAHGIDPNIRTVRT	322				
Db	1225	VAHLHAPTSGSKTKVPAAAYAAQGYKVLVNLPSVAATLFGAYMSKAHGIDPNIRTVRT	1284				
Qy	323	ITTGSPITTYTKFLADGGCSGGAYDIIICDECHSDTATSLIGITVLDQAEATAGARLV	382				
Db	1285	ITTGSPITTYTKFLADGGCSGGAYDIIICDECHSDTATSLIGITVLDQAEATAGARLV	1344				
Qy	383	VLATATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDLA	442				
Db	1345	VLATATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDLA	1404				
Qy	443	AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDSVIDCNTCTVTQVDF	502				
Db	1405	AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDSVIDCNTCTVTQVDF	1464				
Qy	503	SLDPTFTIETITLQDQAVSRTQRRGTGRKPGIYRFVAPGERPSGMFDSVLCYCDAG	562				
Db	1465	SLDPTFTIETITLQDQAVSRTQRRGTGRKPGIYRFVAPGERPSGMFDSVLCYCDAG	1524				

Qy	563	CWYELTPAETTVRLRAYMNTPLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENLP	622
Db	1525	CWYELTPAETTVRLRAYMNTPLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENFP	1584
Qy	623	YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTPLLYRLGAVQNETTLTHPVTKY	682
Db	1585	YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTPLLYRLGAVQNETTLTHPVTKY	1644
Qy	683	IMTCSADLEVVTS-----ACSGKPAIIPDREVLYRE	714
Db	1645	IMTCSADLEVVTSVWLVGGLAALAAAYCLSTGCVVIGRI VLSGRPAIIPDREVLYRE	1704
Qy	715	FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQYVD-----	765
Db	1705	FDEMECSQHLPIYEQGMMLAEQFKQKALGLQATSRQAE-VIAPTQVQTNMCKLEAFWAK	1763
Qy	766	EMECSQAAPVIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEIILYE-----	811
Db	1764	HMWNFIIGIQYLAGLSTLPG--NPAAISLMAFTAAVTSPLTTSQTLLFNILGGWVAQAQLA	1821
Qy	812	-----AFDEMECASKAALIEEGORMAEMLSKIQGLLG-----	845
Db	1822	APGAATFVSGSLAGAAVSGVGLGRVLVDIILAGYGAGVALFAFKIMSGELPSTEDLVN	1881
Qy	846	-----ILRRHVGPGEQAVQWNNRLIAFASRGNHVSPTHYVPS-----	882
Db	1882	LLPAILSPGALVGVCAAILRRHVGPGEQAVQWNNRLIAFASRGNHVSPTHYVPSDAA	1941
Qy	883	-----	882
Db	1942	ARVTAISSLVTQLRLRLHQLWSESSTPCSGSLRLDWDWI CEVLSDFKTLKTKLMP	2001
Qy	883	-----	882
Db	2002	HLGPIPVSCOHGYKGVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPKTCNNMWSGTFF	2061
Qy	883	-----	882
Db	2062	INAVTTGPTCLPAPNYTFALWRVSAEYVEIRRVDPHYVTGTTMDNLKPCQVPSPEF	2121
Qy	883	-----RSRRA-----QALVPWAPDY-----	899
Db	2122	FTELDGVRLHRFAPCKPLLEEVSVFVGLHDYVPVGSQPLCEPEPDVAVLTSMLTDPHSI	2181
Qy	900	-----NRP-----LVET---WKK-----	909
Db	2182	TAAAGRLARGSPSEASASSASQLSAPSLKATCTINHDSPDAELIEANLLWRQEMGNI	2241
Qy	910	-----PDYEPPVPHG-----RSSRFAQALPVWARPDPNPLVETW	945
Db	2242	TRVESENKVILDSFDPLVAEDEREISVPAILKSRRTQALPIWARPDPNPLIETW	2301
Qy	946	KKPDYEPVPVHGKTKRYNRRRQDVKFPGGQIVGRRGPPIPKARR	992
Db	2302	KKPNYEPVPVHGCLP-----PPQ-----SPVPVPPRK	2329
RESULT 3			
GNMVCH			
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)			
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
A:Note: host Homo sapiens (man)			
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004			
C:Accession: A36814; A41546			
R:Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.			
submitted to GenBank, July 1992			
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C			
A:Reference number: A36814			
A:Molecule type: genomic RNA			
A:Residues: 1-3011 <INC>			

A:Cross-references: UNIPROT:P27958; GB:M67463; MID:9329737; PID:AAA45534.1; PID:9329738			
R:Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.			
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991			
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari			
A:Reference number: A41546; MUID:92052256; PMID:1658800			
A:Contents: annotation			
A:Note: neither amino acid nor nucleotide sequence is given			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura			
F:115-Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <EPM>			
F:192-389/Product: major envelope protein E #status predicted <MEE>			
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>			
F:1230-1237/Region: nucleotide-binding motif A (P-loop)			
F:1312-1317/Region: nucleotide-binding motif B			
F:1316-1319/Region: DEXH motif			
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>			
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>			
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23			
Query Match 47.7%; Score 2821; DB 1; Length 3011;			
Best Local Similarity 51.3%; Pred. No. 1.4e-167;			
Matches 611; Conservative 41; Mismatches 110; Indels 430; Gaps 17;			
Qy	204	TGGAARTTSLTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDSSPPVPPQSFQ	262
Db	1165	SGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLETTMRSPVFTDSSPPVPPQSFQ	1224
Qy	263	VAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGDINRTGVRT	322
Db	1225	VAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGVDPNRTGVRT	1284
Qy	323	ITTCSPITYSTYGVFLADGCGSGGAYDIIICDECHSTDATSGILGTVLDQAEAGARLV	382
Db	1285	ITTCSPITYSTYGVFLADGCGSGGAYDIIICDECHSTDATSGILGTVLDQAEAGARLV	1344
Qy	383	VLATATPGSVTVPHNIEEVALSTTGTEIPFYKAIPLEVIKGGRHLLFCHSKKKDELA	442
Db	1345	VLATATPGSVTVSHNIEEVALSTTGTEIPFYKAIPLEVIKGGRHLLFCHSKKKDELA	1404
Qy	443	AKLVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGYTGDPSVDCNTCTVTQVDF	502
Db	1405	AKLVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGYTGDPSVDCNTCTVTQVDF	1464
Qy	503	SLDPTFTTETTLPODAVSRTOGRGRGKPGIYRFVAPGERPSGMFDSVLCBCYDAG	562
Db	1465	SLDPTFTTETTLPODAVSRTOGRGRGKPGIYRFVAPGERPSGMFDSVLCBCYDAG	1524
Qy	563	CWYELTPAETTVRLRAYMNTPLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENLP	622
Db	1525	CWYELTPAETTVRLRAYMNTPLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENFP	1584
Qy	623	YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTPLLYRLGAVQNETTLTHPVTKY	682
Db	1585	YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTPLLYRLGAVQNETTLTHPVTKY	1644
Qy	683	IMTCSADLEVVTS-----ACSGKPAIIPDREVLYRE	714
Db	1645	IMTCSADLEVVTSVWLVGGLAALAAAYCLSTGCVVIGRI VLSGRPAIIPDREVLYRE	1704
Qy	715	FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQYDEMERC	770
Db	1705	FDEMECSQHLPIYEQGMMLAEQFKQKALGLQATSRHAE-VITPAVQTNMCKL-EVFWA	1762
Qy	771	SOAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP-----VTPDKKEIILYE-----	811
Db	1763	KHMWNFIIGIQYLAG-----LSTPGNPAISLMAFTAAVTSPLTGTGTLFNILGGWV	1816
Qy	812	-----AFDEMECASKAALIEEGORMAEMLSKIQGLLG-----	845



Db	2214	TCETHSDPADLIEANLLMRQMGNIITRVESKNVVLDSFDPLRAEEDERSVAAE	2273
Qy	920	--RSRRFAQALPVYKARDYNNPLVETWKKDDYPPVHVHGRKTKRNTNRRPDQVKFPGG	977
Db	2274	ILRSKFPFPPALPIWARDYNNPLLESKSDYVPPAVHGCPLPPTT-----	2320
Qy	978	QIVGRRGPPIPKARR	992
Db	2321	-----GPIPPPRK	2329
RESULT 5			
GNWVCJ			
Genome polyprotein - hepatitis C virus (strain J)			
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004			
C:Accession: A39253; PS0086			
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto			
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990			
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients			
A:Reference number: A39253; MUID:91088550; PMID:2175903			
A:Accession: A39253			
A:Molecule type: genomic RNA			
A:Residues: 1-3010 <KAT>			
A:Cross-references: UNIPROT:P26662; GB:D90208; NID:G221610; PIDN:BAAL14233.1; PID:G221611			
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.			
Proc. Jpn. Acad. 65B, 219-223, 1989			
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari			
A:Reference number: PS0085			
A:Accession: PS0086			
A:Molecule type: genomic RNA			
A:Residues: 2650-2707 <KAT>			
A:Experimental source: Japanese isolate			
C:Comment: The cleavage sites of this polyprotein have not been determined.			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin			
F:2-115/Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <EPM>			
F:192-389/Product: major envelope protein E #status predicted <MEE>			
F:730-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>			
F:1230-1237/Region: nucleotide-binding motif A (P-loop)			
F:1312-1317/Region: nucleotide-binding motif B			
F:1316-1319/Region: DEXH motif			
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>			
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>			
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2			
Query Match 46.4%; Score 2746; DB 1; Length 3010;			
Best Local Similarity 48.2%; Pred. No. 6.7e-163;			
Matches 579; Conservative 58; Mismatches 116; Indels 448; Gaps 14;			
Qy	204	TGGAARTSGLTSLFSPG-ASQNIQLIVDFIPVENLETTMESPVFTDNSSPPVVFQSFQ	262
Db	1165	SGGPLLCPGHHVVGIFRAVCTRGVAKAVDFIPVSMETTRMSPVFTDNSSPPVVFQSFQ	1224
Qy	263	VAHLHAPTSKSTKVPAYAAQGYKVLNPSVAATLFGAYMSKAHGDINRTGVRT	322
Db	1225	VAHLHAPTSKSTKVPAYAAQGYKVLNPSVAATLFGAYMSKAHGDINRTGVRT	1284
Qy	323	ITTGSPITVSTYKFLADGCGSGGAYDIIICDECHSTADTSILGTVLDDQAEAGARLV	382
Db	1285	ITTGSPITVSTYKFLADGCGSGGAYDIIICDECHSTADTSILGTVLDDQAEAGARLV	1344
Qy	383	VLATATPPGSVTVPHNIEEVALSTGEIPFYKAIPLVIVKGGHLLIFCHSKKKDELA	442
Db	1345	VLATATPPGSVTVPHNIEEVALSTGEIPFYKAIPLVIVKGGHLLIFCHSKKKDELA	1404

RESULT 6

GNWVC

Genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural

Qy	443	AKLVALGINAVAYYRGDVSVIPTSGDVVVVATDALTMTGYTGDSDSDVIDCNTCTVTQTVDF	502
Db	1405	AKLTGLGLNNAVAYYRGDVSVIPTSGDVVVVATDALTMTGYTGDSDSDVIDCNTCTVTQTVDF	1464
Qy	503	SLDPTFTIETITLPODAVSRTRGRGKPGIYRFVAPGPERPSGMFSDSSVLCECYDAG	562
Db	1465	SLDPTFTIETITLPODAVSRTRGRGKPGIYRFVAPGPERPSGMFSDSSVLCECYDAG	1524
Qy	563	CWYELTPAETTVRLRAYMNTPLPVQODHLEFEGVFTGLTHDAHFLSOTKSGENLP	622
Db	1525	CWYELTPAETTVRLRAYMNTPLPVQODHLEFEGVFTGLTHDAHFLSOTKSGENLP	1584
Qy	623	YLVAQATVCAACAQPPSWDMWKLRLKPTLHPTPLLYRLCAVONEITLTHPTVKY	682
Db	1585	YLVAQATVCAACAQPPSWDMWKLRLKPTLHPTPLLYRLCAVONEITLTHPTVKY	1644
Qy	683	IMTQMSADLEVVVTS- - - - -CSGKPAIIPDREVLVRE	714
Db	1645	IMTQMSADLEVVVTS- - - - -CSGKPAIIPDREVLVRE	1704
Qy	715	PDMEECASHLPYIEQGMMLAEQKQKALGSLRGKPAIVDPKEVLYQYDEMEECSQAA	774
Db	1705	PDMEECASHLPYIEQGMMLAEQKQKALGSLRGKPAIVDPKEVLYQYDEMEECSQAA	1747
Qy	775	PIIEQAQ- - - - -VIAHQFKEKVLGLINDQVV- - - - -VTPDKEI	808
Db	1748	PIIEQAQ- - - - -VIAHQFKEKVLGLINDQVV- - - - -VTPDKEI	1807
Qy	809	LY- - - - -EAFDEMEECASKAALIEGQMAEMLKSIQGLG- - - - -	845
Db	1808	LY- - - - -EAFDEMEECASKAALIEGQMAEMLKSIQGLG- - - - -	1867
Qy	846	- - - - -ILRRHVGPGEQAVQMMRLIAPASRGN	872
Db	1868	- - - - -ILRRHVGPGEQAVQMMRLIAPASRGN	1927
Qy	873	HVSTHYVP- - - - -SRSRFAQALPV- - - - -WAPDYNPLVETWKK- - - - -	909
Db	1928	HVSTHYVP- - - - -SRSRFAQALPV- - - - -WAPDYNPLVETWKK- - - - -	1987
Qy	910	- - - - -PDY- - - - -	909
Db	1988	- - - - -PDY- - - - -	2047
Qy	910	- - - - -PDY- - - - -	912
Db	2048	- - - - -PDY- - - - -	2107
Qy	913	- - - - -BPP	915
Db	2108	- - - - -BPP	2167
Qy	916	- - - - -HG- - - - -	919
Db	2168	- - - - -HG- - - - -	2227
Qy	920	- - - - -RSSRFAQALPV	931
Db	2228	- - - - -RSSRFAQALPV	2287
Qy	932	WAPDYNPLVETWKKPDYEPVHVHGRKTKRNTNRRPDQVKFPGGQIVVGRGPPIPKAR	991
Db	2288	WAPDYNPLVETWKKPDYEPVHVHGRKTKRNTNRRPDQVKFPGGQIVVGRGPPIPKAR	2328
Qy	992	R	992
Db	2329	R	2329



protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: A38465  
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;  
J. Virol. 65, 1105-1113, 1991  
A;Title: Structure and organization of the hepatitis C virus genome isolated from human  
A;Reference number: A38465; MUID:91140698; PMID:1847440  
A;Accession: A38465  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <TAK>  
A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g3297  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 46.4%; Score 2741.5; DB 1; Length 3010;  
Best Local Similarity 48.9%; Pred. No. 1.3e-162;  
Matches 577; Conservative 56; Mismatches 100; Indels 447; Gaps 14;

QY 224 SONQLIVDFIPVENLETTMRSPVFTDNSSPVVPSQFQVAHLHAPTSKSTKVPAAAYA 283  
DB 1186 TRGAKAVDFPVVESMTMRSPVFTDNSSPVVPSQFQVAHLHAPTSKSTKVPAAAYA 1245  
QY 284 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTGRTITTTGSPITYSTYGRFLADGGC 343  
DB 1246 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTGRTITTTGAPVTYSTYGRFLADGGC 1305  
QY 344 SGAYDIIICDECHSTDATILGIGTVDLQAEAGARLVVLATATPPGSTVTPHPNIEEV 403  
DB 1306 SGAYDIIICDECHSTDATILGIGTVDLQAEAGARLVVLATATPPGSTVTPHPNIEEV 1365  
QY 404 ALSTTGTIPYGYKAIPUEVTKGRHLIFCHSKKCDLAAKLVAGLNNAVAYRGLDVSV 463  
DB 1366 ALSTTGTIPYGYKAIPUEVTKGRHLIFCHSKKCDLAAKLVAGLNNAVAYRGLDVSV 1425  
QY 464 IPTSGDVVVVATDALMTGYTGDFSDVIDCNTCTVTQTVDLDFPTFTTITLTPQDAVSRT 523  
DB 1426 IPTSGDVVVVATDALMTGYTGDFSDVIDCNTCTVTQTVDLDFPTFTTITLTPQDAVSRS 1485  
QY 524 QRRGRTGRGPIYRFVAPGRPSGMDFSSVLCYCYDAGCAWYBELTTPAETTVRLAYMNT 583  
DB 1486 QRRGRTGRGPIYRFVAPGRPSGMDFSSVLCYCYDAGCAWYBELTTPAETTVRLAYMNT 1545  
QY 584 PGLPVCODHLEFVEGVTGLTHIDAHFLSOTKSGENLPYLVAQVATVCARAQAPPPSWD 643  
DB 1546 PGLPVCODHLEFVEGVTGLTHIDAHFLSOTKSGENLPYLVAQVATVCARAQAPPPSWD 1605  
QY 644 QMWKCLIRLPTLHGPTPLLYRLGAVQNEITLHPVTKYIMTCSADLEVVYTSVSA----- 697  
DB 1606 QMWKCLIRLPTLHGPTPLLYRLGAVQNEITLHPVTKYIMTCSADLEVVYTSVSA----- 1665  
QY 698 -----CSGKPAIIPDRVLVRFDEMECECSQHLPYIEQGMMLA 735  
DB 1666 VLAALAAAYCLTGSWIVGRILSGRAIPVDRRELLYQEFDEMECEASHLPYIEQGMMLA 1725  
QY 736 EQFKQKALGSRGKPAIPVDPKEVLYQYDEMESCQAAPIEQAOVIAHQFEK-VLGL 794  
DB 1726 EQFKQKALGLUQ-----TATKQAAAPVVEKSWKRALETFWAKGHWNF 1768  
QY 795 IDNDQVV-----VTPDKELLY-----EA 812

Db 1769 ISGIQYLAGLSTLPGNPAIASLMAFTASITSTLTTQSTLLFNILGWMVAQAAPPASA 1828  
QY 813 FDEMEECASKAALIEBGGQMAEMLKSKIOGLLQ----- 845  
Db 1829 FVAGIAGAAVSGSIGLUGKLVLDILAGYGAGVAGALVAFKVMGSEMPSTEDLVNLLPAIL 1888  
QY 846 -----ILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPS----- 882  
Db 1889 PGALVVGVCAAILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPSDDAAARVTQIL 1948  
QY 883 ----- 882  
Db 1949 SSLITQTLKRLHQMINECDSTPCSGSLRDVMDWICTVLTDFTKTLQSKLLPOLPGVPF 2008  
QY 883 ----- 882  
Db 2009 FSCQGYKGVWRDGIQMOTPCGAQITGHVKNGSMRIVGPKTCSNTWHTGTFPINAYTTG 2068  
QY 883 ----- 882  
Db 2069 PCTPSPAPNYSRALMRVAEEVETRVGDFHYVTGTTDNVKKPCQVPAPEFSEVDGV 2128  
QY 883 RSRPFAQA-----LPVWARPDY----- 899  
Db 2129 RLHRYAPACRPLLRBEVTRFQVGLNQVLGSQLPCEPEPDVAVLTSLMTPDPSHITAETAKR 2188  
QY 900 -----NPP-----LVET---WKK----- 909  
Db 2189 RLARGSPFSLASSASSQLSAPLSKATCTTHVHSPDADLLEANLLWRQMGNGNTRVSEEN 2248  
QY 910 -----PDYEP-----PVVHGRSSRRFAQALPVWARPDPNPPVETWKKPDYEP 952  
Db 2249 KVVVLDSPDLRAEEDEREVSVAETLRKSKKFAAPMIAWARPDPNPPLESWKDPDYVP 2308  
QY 953 PVVHGRKTRNTNRRPDQVKFGGQIVGRRGPPPIKARR 992  
Db 2309 PVVH-----GCPLPPKAPPPIPPRR 2329

RESULT 7  
S18030  
genome polyprotein - hepatitis C virus (isolate JK1)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus NS4a; nonstructural protein NS5  
C;Species: hepatitis C virus  
A;Variety: isolate JK1  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: S18030; S33570; A48332; S18029  
R;Honda, M.; Kaneko, S.; Masahashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
A;Reference number: S18028  
A;Accession: S18030  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <HON>  
A;Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
A;Experimental source: isolate JK1 from an individual  
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated from  
A;Reference number: A48332; MUID:93119270; PMID:8380322  
A;Accession: S33570  
A;Molecule type: genomic RNA  
A;Residues: 1-547; 'T', 549-621; 'V', 623-624; 'S', 626-652; 'DL', 655-761; 'T', 763-782 <HON>  
A;Cross-references: EMBL:X61591  
A;Note: this sequence is inconsistent with the nucleotide translation  
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as  
as Trp, and TTC for residue 771 as Ser  
A;Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F;2-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product:	envelope protein M	#status predicted <EPM>
F;192-389/Product:	major envelope protein E	#status predicted <MEE>
F;390-729/Product:	nonstructural protein NS1	#status predicted <NS1>
F;730-1006/Product:	nonstructural protein NS2	#status predicted <NS2>
F;1007-1615/Product:	hepacivirin	#status predicted <NS3>
F;1230-1237/Region:	nucleotide-binding motif A (P-loop)	
F;1312-1317/Region:	nucleotide-binding motif B	
F;1316-1319/Region:	DEXH motif	
F;1616-1862/Product:	nonstructural protein NS4a	#status predicted <N4a>
F;1863-2013/Product:	nonstructural protein NS4b	#status predicted <N4b>
F;2014-3010/Product:	nonstructural protein NS5	#status predicted <NS5>
F;196,209,234,250,305,417,423,448,536,540,556,576,623,645/Binding site:	carbohydrate (Asn)	

  

Query Match	46.3%;	Score	2737.5;	DB 1;	Length	3010;			
Best Local Similarity	46.4%;	Pred. No.	2.3e-162;						
Matches	599;	Conservative	55;	Mismatches	140;	Indels	497;	Gaps	18;

  

Qy	190	SLPAPGAKQNETHVTGGAAARTTSGLTSLFSPG-A	ASNTQLIYVDPIPVENLETTMRSPVP	248
Db	1151	SLLSPPRVSYLKGSSGPLLCPSGHAVGIFRAAIVCTRGVAKAVDFIPVESMETTRSPVF	1210	
Qy	249	TDNSSPPVPOSFOVAHLHAPTCGSKSTKVPAAAYAGQYKVLVNLNPSVAATLFGGAYMSK	308	
Db	1211	TDNSSPPAVPQTQVAHLHAPTCGSKSTKVPAAAYAGQYKVLVNLNPSVAATLFGGAYMSK	1270	
Qy	309	AHGIDPNIRGTITTTGSPITYTYTGKFLADGGCSGGAYDIIICDECHSTDATSILGIG	368	
Db	1271	AHGVDPNISVGRTITTYTGAITTYTYTGKFLADGGCSGGAYDIIICDECHSTDSTSILGIG	1330	
Qy	369	TVLDQAETAGARLVLAATATPPGSVTPVPHNIEEVALSTTGEIPFYKAIPLVEIKGGRH	428	
Db	1331	TVLDQAETAGARLVLAATAATPPGSVTPVPHNIEEVALPNTGEIPFYKAIPLVEIKGGRH	1390	
Qy	429	LIFCHSKKKODELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFFS	488	
Db	1391	LIFCHSKKKODELAALKSALGNANAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFFS	1450	
Qy	489	VIDCNTCVTCVTFDFSLDPTFTIETILPDQAVSRKTORRGRTGKPGIYRFVAPGERPSG	548	
Db	1451	VIDCNTCVTCVTFDFSLDPTFTIETTLPDQAVSRQRGRHTGRRGIYRFVTPGERPSG	1510	
Qy	549	MFDSVLCCECYDAGCAWYELTPAETTIVRLRAYNMNTPGLPVCODHLEFWEGVFTGLTHIDA	608	
Db	1511	MFDSVLCCECYDAGCAWYELTPAVTSVRLRAYLNTPGLPVCQHLEFWESVFTGLTHIDA	1570	
Qy	609	HFLSQTKQSENIPLYVAYQATVCARAQAAPPWDQWKCLIRLKPTLHGPTLLYRLGA	668	
Db	1571	HFLSQTKQAGENPYLVAYQATVCARAQAAPPWDQWKCLIRLKPTLHGPTLLYRLGA	1630	
Qy	669	VQNEITLTHPVTKYIMTCMSADLEVVTSA-----CSG	700	
Db	1631	VQNEVTLTHPTTPIMACMSADLEVVTSTWVLVGVLAALAAAYCLTTGVSVVIVGRILLSG	1690	
Qy	701	KPAIIPREVLYREFDEMEECSOHLPIYEOQMMLAEQFKOKALGLSRGGKPAIVDPKEVL	760	
Db	1691	RPAIIPREVLYQEFDEMEECASHLPIYEQMQLAQEFKQKALLGTTAS-----	1740	
Qy	761	YYQYDEMEECSPAAYIE-----QAQVI AH-----QFKEKVLGLINDQVV-----	801	
Db	1741	-----KQAEAAAAPVYESKQWALEAFWAXHMWNFISGIQVYLAGSLTLPGNPAIVSLMAF	1793	
Qy	802	-----VTPDKELY-----EAPDEMEECASKAALIEEGQMAEMLK	837	
Db	1794	TASITSPLTTOHTILLFILGWAAQAPPASAASFVAGAGTAGAAGVSI GLGKLVVDILA	1853	
Qy	838	SKI OGLLG-----ILRRHVGPGEVAV	858	
Db	1854	GYGAGVAGALVAFKMGGMPESTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEVAV	1913	
Qy	859	QWMNRLIAFASRGNHVSPTHYP-----SRSRREFAQALPVWARPDYN	900	
Db	1914	QWMNRLIAFASRGNHVSPTHVVPESDAARVTKILLSITITQRLRLHQ-----WIINDCS	1969	

[illegible]

QY 263 VAHLHAPTCGSKSTKVPAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGIDNIRTVGT 322  
DB 1225 VAHLHAPTCGSKSTKVPAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGIDNIRTVGT 1284  
QY 323 ITTGSPTITYTGKFLADGGCGGAYDIIICDCHSTDATSIIGTGLVLDQAETAGARLV 382  
DB 1285 ITTGAPITTYTGKFLADGGCGGAYDIIICDCHSTDTSTIIGTGLVLDQAETAGARLV 1344  
QY 383 VLATATPPGSVTVPHPNIEBEVALSTGTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 442  
DB 1345 VLATATPPGSVTVPHPNIEBEVALSTGTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 1404  
QY 443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDVSIDNCVTCVTVDF 502  
DB 1405 AKLSALGINAVAYRGLDVSIVPASGNVVVATDALMTGTFGDFDVSIDNCVTCVTVDF 1464  
QY 503 SLDPPTTITITLPODAVSRTRGRGTGKPGIYRFVAPGERPSGMFSSVLCYCYDAG 562  
DB 1465 SLDPPTTITITWPDQAVSRQRGRTSRGRGIYRFVTPGERPSGMFSSVLCYCYDAG 1524  
QY 563 CAWYELTPAETTVRLRAYMNTPLGPVQCQDHFLEFWGFTGLTHIDAHFLSQTKQSGENLP 622  
DB 1525 CAWYELTPAETSVRLRAYLNTPLGPVQCQDHFLEFWGFTGLTHIDAHFLSQTKQAGDNFP 1584  
QY 623 YLVAYQATVCARAQAPPSWDQWKKLIRLKPTLHGPTPLLYRLGAVONEITLTHPTVKY 682  
DB 1585 YLVAYQATVCARAQAPPSWDQWKKLIRLKPTLHGPTPLLYRLGAVONEITLTHPTVKY 1644  
QY 683 IMTCMSADLEWTSIA-----CSGKPAIIPDREVLRYRE 714  
DB 1645 IMACMSADLEWTSITWVLVGGVLAALAAVCLTTGVSIVVIGRIILSKPAPVPDREVLRYOE 1704  
QY 715 FDEMECSOHLPIYEQGMMLAEQFKQALGLSRGKPAIVPDKEVLYQQYDEMECSQAA 774  
DB 1705 FDEMEECASHPLPIYEQGMQOLAEQFKQALGLLQ-----TATKQAEAAA 1747  
QY 775 PYIEQAO-----VTAHQFEKVKGLINDQVV-----VTPDKEI 808  
DB 1748 PVYESKRWLTLEAFWANDMNFISIGIYLAGLSTLPGNPAIASLMAFTASITSLTQSTL 1807  
QY 809 LY-----EAFDEMECASKAALIEEGORMASMLKSKQGLLG-----845  
DB 1808 LFNILGGVNAQAQLAPGAASAFVAGIAGNAGVCSIGLKGKVLVDNVAGYAGVAGALVAFK 1867  
QY 846 -----ILRRHVGPGEVQVGMNRLIAFASRGN 872  
DB 1868 VMSGEMPSTEDLVNLLPAILSPCALVGVVCAAILRRHVDPGEGAVQVMNRLIAFASRGN 1927  
QY 873 HVSPTHYVPS-----882  
DB 1928 HVSPTHYVPSDAAARVTOILSGLTITQLLRLHQMINEPCSTPCSGSMLRDVMDWICTV 1987  
QY 883 -----882  
DB 1988 LADFKTWLQSLPLRPLPGVPFFSCQRGYKGVWRGDGIMQTTCPGQAQLTGHVKNXGSMRIW 2047  
QY 883 -----882  
DB 2048 GPKTCSNTWHGTFFPINAYTTGCTPSPAPNYSRALMRVAEEYVEVRVGDHFVYTGMTT 2107  
QY 883 -----RSRRFAQA-----LPVWARPD 898  
DB 2108 DNVKPCQVPAPBFFTEVDGVRHLRYAPACKPLLRBEVSPQVGLNOYVWGSQLPCEPEPD 2167  
QY 899 -----YNPP-----LV 904  
DB 2168 VAVLTSMLTDPSSHITAETAKRRLARGSPPSLASSASQLSALSLSKAACTRTHRTPPADLI 2227  
QY 905 ET---WKK-----PDYEP-----PVHGSSRRFAQALPV 931  
DB 2228 EANLLWRQMGGINTRVESENKVVILDSFDPLRAEEDEREVSVPABILLRKSRRKFPFPAALPV 2287

QY 932 WARPDYNPPLVETWKKPDYEPVPHVGRKTKRNTNRRPQDVKPPGGQIQVGRGPPPIPKAR 991  
DB 2288 WARPDYNPPLLEPWKDPDPVPPVPH-----CCPLPPVKAPPIPPR 2328  
QY 992 R 992  
DB 2329 R 2329  
RESULT 9  
JC5620  
genome polyprotein - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: JC5620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant s  
A:Reference number: JC5620; MUID:97366593; PMID:9223423  
A:Accession: JC5620  
A:Molecule type: mRNA  
A:Residues: 1-3014 <CHA>  
A:Cross-references: UNIPROT:O39928; GB:Y13184  
A:Experimental source: genotype 5a, which predominates in South Africa  
A>Note: the translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:384-408/Region: hypervariable #status predicted  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: hepatitisvirin #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NA>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NB>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interferon sensitivity determining #status predicted  
Query Match 41.5%; Score 2455.5; DB 1; Length 3014;  
Best Local Similarity 42.9%; Pred. No. 18-144;  
Matches 531; Conservative 85; Mismatches 172; Indels 450; Gaps 17;  
QY 179 GSAARTTSGFVSILF-----APGAKQNETHV-----TCGAAARTTSGLT 216  
DB 1119 GSLTRCTCGSADLYLVTRHADVIPARRGDTTRASLLSPRISVLKSGSGPIWCPSGHV 1178  
QY 217 SLFSPG-ASONTQILIVDFIPVENLETTMRSPVFTDNSSPPVPPQSQVAHLHAPTSGSKS 275  
DB 1179 GVFRAAVCTRGVAKALEFVVENLETTMRSPVFTDNSTPPAVPHEFQVGHLPAPTSGSKS 1238  
QY 276 TKVPAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGIDNIRTVGTITGSPITYSTVG 335  
DB 1239 TKVPAAYAAQGYKVLVLPNSVAATLFGAYMSRAYGVDPNIRTVGTITGAGITYSTYG 1298  
QY 336 KFLADGGCGGAYDIIICDCHSTDATSIIGTGLVLDQAETAGARLVVLATATPPGSVTV 395  
DB 1299 KFFADGGCGGAYDVIICDCHSQDATTILGTVLDQAETAGARLVVLATATPPGSVIT 1358  
QY 396 PHPNIEBEVALSTGTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELAALVALGINAVAY 455  
DB 1359 PHPNIEBEVALPSEGEIPFYGRAIPLVLIKGRHLIFANOKKAKETAKONKPGREAVAY 1418  
QY 456 YRGLDVSIVPTSGDVVVVATDALMTGYTGDFDVSIDNCVTCVTVDFSLDPTFTITL 515  
DB 1419 YRGLDVAVIPATGDVVVVCSTDALMTGTFGDFDVSIDNCNSAVTQTVDFSLDPTFTITTV 1478  
QY 516 PQDAVSRTRGRGTGKPGIYRFVAPGERPSGMFSSVLCYCYDAGCAWYELTPAETTV 575

Db	1479	POBAVSQRGRGTRGRHGIIYRVSSGERPSGIFDSVLCECYDAGCAWYDITPAETTV	1538
Qy	576	RLRAYMTGCLPVCQDHLEFWEGVFTGLTHIDAHFLSQTQSGENLPYLVAIQATVCARA	635
Db	1539	RLRAYLMTGCLPVCQDHLEFWEGVFTGLTHIDAHFLSQTQSGENLPYLVAIQATVCARA	1598
Qy	636	QAPPSQDMWKCLIRLKTPLHGTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVT	695
Db	1599	KAPPSQDMWKCLIRLKTPLHGTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVT	1658
Qy	696	SA-----CSGPAIIIPOREVLYRFEDEMECSQHLPY	727
Db	1659	STWLVGSGVAALAAAYCLTWGSVAIVGRILLSGRPAITPDREVLYQOFDEMECSALPY	1718
Qy	728	IEQGMMLAEQFKQKALGL--SRGKPAIVDPKEVLYQOYDEMEB-----	769
Db	1719	VDEARATAGQKEKVLGLIGTAGQKAETL--KPAATSMWSKAEGFWAKHMMNFVSGIOYL	1776
Qy	770	-----CSQAAPYIEQAQVIAHQFK	788
Db	1777	AGLSTLPGNPAVATLMSFTAATVTSPLTTHOTLLFNILGGMVASQIAPPTAATAFVSGMA	1836
Qy	789	EKVLGLINDOVVTPPKXELIYEAFFDEMEECASKAALIEGQR--MAEMLKSKIQGLL---	844
Db	1837	GAAVGNIGLGRVLID-----ILAGYGTGVAGALVAFKIMCGERTPTABELVNLPSILCPG	1891
Qy	845	-----GILRRHVGEGAVQNMNRLIAFASRGNHVSPTHVVP--SRSRRAQALP-	892
Db	1892	ALVGVGTCAAVLRRHIGEGAVQNMNRLIAFASRGNHGSPTHVVPETDASAKVTQLLSS	1951
Qy	893	-----VMARPDYNPLVETWK-	909
Db	1952	LTVTSLLKRLHTWIGEDYSTPCDGTWLRAIWDWVCTALTDFKAWLQAKLLPOLPGVPFFS	2011
Qy	910	-----	909
Db	2012	CQKGYKGYWRGDGVNSTKPCGATISGHVKNGTMRIVGPKLCSNTWQGTFFINATTTGPS	2071
Qy	910	-----PDYE-----	913
Db	2072	VPAPAPNTKFWLRVGAADYAEVRVRGDYHITGVTDQNLKPCQVSPSEFFTELDGVRI	2131
Qy	914	-----	913
Db	2132	HRFAPPNPLLREEVTFVSLHSYVGSQLPCEPEPDVTVLTSMLSDPAHITAETAKERL	2191
Qy	914	-----PPVHGRSSRR-----	924
Db	2192	NRGSPPSLANSSASQLSAPSLKATCTIQHHPDADLIKANILLWRQCMGNITRVEAENKV	2251
Qy	925	-----FAQALPVWARPDPYNPPLVETWKPKDPYBPPV	954
Db	2252	EILDCFPLKEEDDRETSVADCFKGPAPFPALPVWARPDPYBPPVETWKPKDPYBPPV	2311
Qy	955	VHGRKTKRNTNRPPQDVAFPGGQIVGRRGPPPKARR	992
Db	2312	WVG-----CFIPNAGR-----PPVPLPRR	2330
RESULT 10			
JQ1303			
genome polyprotein - hepatitis C virus (isolate HC-96)			
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate HC-96) nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004			
C:Accession: JQ1303			
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.			
J. Gen. Virol. 72, 2697-2704, 1991			
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human			
A:Reference number: JQ1303; MUID:92044440; PMID:1658196			
A:Accession: JQ1303			

A:Molecule type: genomic RNA			
A:Residues: 1-3033 <OKA>			
A:Cross-references: UNIPROT:P26660; GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651			
C:Experimental source: isolate HC-J6 from a Japanese individual			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane			
F:2-115/Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <BPM>			
F:192-389/Product: major envelope protein E #status predicted <HEE>			
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>			
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1011-1615/Product: hepatitis C virus nonstructural protein NS3 #status predicted <NS3>			
F:1316-1321/Region: DEXH motif			
F:1320-1323/Region: DEXH motif			
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>			
F:1867-2017/Product: nonstructural protein NS5 #status predicted <N4b>			
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NOS>			
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 477, 534, 542, 558, 578, 627, 649, 1091, 1217, 1259, 2038, 21			
Query Match 41.1%; Score 2431; DB 1; Length 3033;			
Best Local Similarity 41.0%; Pred. No. 3.4e-143;			
Matches 547; Conservative 61; Mismatches 125; Indels 602; Gaps 20;			
Qy	224	SONIQLVDFIPVENLETTMRSPVFTDNSSPPVQSFQVAHLHAPTGSKSTKVPAAVA	283
Db	1190	SRGAKSIDIPVETLIDIVTRSPFSDNSPPAVPQYQVGLHAPTGSKSTKVPAAVA	1249
Qy	284	AQGYKVLVLPNSVAATLGFAGYMSKAHGDINIRGTVRTITTTGSPITVSTYKFLADGGC	343
Db	1250	AQGYKVLVLPNSVAATLGFAGYMSKAHGINPRTVTRTITTTGAPITVSTYKFLADGGC	1309
Qy	344	SGGAYDIIICDECHSTATSIIGTGLDQETAGARLVLATATPPGCVTPVHPNIEV	403
Db	1310	AGGAYDIIICDECHAVDSTTILGTVLDQETAGARLVLATATPPGCVTPVHPNIEV	1369
Qy	404	ALSTTGEIPFYGKAIPLVFKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS	463
Db	1370	ALGEGEIPFYGKAIPLVFKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVS	1429
Qy	464	IPTSGDWWATDALMTGYTGDPSVIDCNTCVTQVDFSLDPTFTTITLTPQDAVSR	523
Db	1430	IPTQGDWWATDALMTGYTGDPSVIDCNVAVTQVDFSLDPTFTTITLTPQDAVSR	1489
Qy	524	QRRGRTGRGKGIYRVFVAPGERPSGMDSSVLCYCYDAGCAWYELTPAETTVRLRAYNT	583
Db	1490	QRRGRTGRGKGIYRVFVAPGERPSGMDSSVLCYCYDAGCAWYELTPAETTVRLRAYNT	1549
Qy	584	PGLPVCODHLEFEGVFTGLTHIDAHFLSQTQSGENLPYLVAIQATVCARAQAPPSWD	643
Db	1550	PGLPVCODHLEFEGVFTGLTHIDAHFLSQTQSGENLPYLVAIQATVCARAQAPPSWD	1609
Qy	644	QMWKCLIRLKTPLHGTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVTSACSGKPA	703
Db	1610	QMWKCLIRLKTPLHGTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVTSACSGKPA	1663
Qy	704	IIPDREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGLSRGCKPAIVDPKEVLVQ	763
Db	1664	IIPDREVLYREFDEMECSQHLPIYEQGMMLAEQFKQKALGLSRGCKPAIVDPKEVLVQ	1672
Qy	764	YDEMECSQAAPYIEQAQVIAHQFKKEKVLGLID-NDQVVVTPDKKEILYEAFDEMEECASK	822
Db	1673	YDEMECSQAAPYIEQAQVIAHQFKKEKVLGLID-NDQVVVTPDKKEILYEAFDEMEECASK	1718
Qy	823	AALIEGORMAEMLSKTIQGLL-----	844
Db	1719	AALIEGORMAEMLSKTIQGLL-----	1778
Qy	845	-----	844
Db	1779	LAGLSTLPGNPAVAMMAFSAALTSPLSTTILLNLLGGLASQIAPPAGATGFVWSGL	1838
Qy	845	-----	844

Db 1839 VGAAGVSGIGKVLVDILACYGAGISGALVAFKIMSGKPSMEDVNVNLLPGLISPGALVV 1898  
QY 845 -----GILRRHVGPGEAGVQWNRLIAFASRGHNVSTHYV--PSRSRRFAQAL----- 891  
Db 1899 GVICAALLRHVGPGEAGVQWNRLIAFASRGHNVSTHYVTHVTSADASQRTVQLGSLTIT 1958  
QY 892 ----- 891  
Db 1959 SLLRRLHNWITDCPIPCSGSLRDLVMDVCTILTDFKNWLTSLKPLPKMPGLPFISQKG 2018  
QY 892 --PWNA----- 895  
Db 2019 YKGVNAGTGMTTRCPGCGANISGNVRLGSMRITGPKTCMNWQGTFFPINCYTEGCQVCPK 2078  
QY 896 RPDVNPPLVETWK----- 908  
Db 2079 APNFK---IAIWRVAASEYAEVTHQGSYHYITGLTTLNKLVPQLSPFESFVWDGVQIH 2135  
QY 909 -----KP-----DYPE----- 914  
Db 2136 RFAPIPKPPRDEVSCVGLNSFVGSQPCDPEPDTDLTSMLTDPSHITAETAAARLLA 2195  
QY 915 ----- 914  
Db 2196 RGSPPSEASSASQLSAPSLRATCTTHGKAYDMDVMDANLFMGQDVTRIESESKVNVLDS 2255  
QY 915 --PVVHGRS-----SRRPAQALPVNARPDYNPLVETWKKPDYEPVHVH-- 957  
Db 2256 LDPWVERSLDERSIPEYMLPKRPPALPAWARPDPYNPLVESWKRPDYQATVAGCA 2315  
QY 958 ----RKTNRNRRPDQVKF-----PGGQGV-----GRR 983  
Db 2316 LPPPKXTPTPPRRRTVGLSESIADALQALIKSFGQPPPSGDSGLSTGADAADSGSR 2375  
QY 984 GPPIPKARRPEGTWAGQVPWPLVGNKRRSTGKSGKPGYWPWRKTKNTNRRPDQVK 1043  
Db 2376 TPDDELALSTGSISSMP-----PLE-----GEPGDP-----DLEPEQVE 2410  
QY 1044 F---PGGQGVVGRGG 1055  
Db 2411 LQPPPGGVVTPGSG 2425

RESULT 11  
GNMVJ8  
genome polyprotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J8) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A40250; PQ0397; PQ0559  
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to reported reference number: A40250; MUID:92230232; PMID:1314459  
A:Reference number: A40250  
A:Accession: A40250  
A:Molecule type: Genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:9221608; PIDN:BAA01761.1; R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to other hepatitis C virus types  
A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0397  
A:Molecule type: Genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D10134  
A:Experimental source: isolate E-b12  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno, Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: PQ0554; MUID:92068204; PMID:1720309  
A:Accession: PQ0559

A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:9221523; PIDN:BAA01418.1; PID:9221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein NS1 #status predicted <NS1>  
F:390-733/Product: nonstructural protein NS2 #status predicted <NS2>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS3>  
F:1011-1639/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS3>  
F:1234-1241/Region: nucleotide-binding motif A (P-loop)  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196-209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,231;  
Query Match 41.1%; Score 2428; DB 1; Length 3033;  
Best Local Similarity 41.9%; Pred. No. 5.3e-143;  
Matches 543; Conservative 54; Mismatches 128; Indels 570; Gaps 18;  
QY 204 TGAAGATTGSLTSLFSPG-ASQNIQLIVDFIPVENLETMRSPVFTDSSPPVPPQSFQ 262  
Db 1169 SGGPVLCSRGHAGVGLFRAAFCARGVAKSIDFIPVESLDVATRTPSFSDNSTPPAVPQSYQ 1228  
QY 263 VAHLHAPTGGSKTKVPAAYAAQYKVLVNSVAATLGFAYMSKAHGDIPNIRTVRT 322  
Db 1229 VGLHAPTGGSKTKVPAAYAAQYKVLVNSVAATLGFAYMSKAHGINPNIIRTVRT 1288  
QY 323 ITTGSPTTSTYTKFLADGGCGGAYDIIICDEHSTDATSLICIGTVLDDQAEATAGARLV 382  
Db 1289 VTGDSITTYTKFLADGGCGGAYDIIICDEHSTDATSLICIGTVLDDQAEATAGARLV 1348  
QY 383 VLATATPPGTVTPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA 442  
Db 1349 VLATATPPGTVTPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA 1408  
QY 443 AKLVAGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVDCNTCTVTVDF 502  
Db 1409 AALRGVNAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVDCNTCTVTVDF 1468  
QY 503 SLDPFTTITITLPQDAVSRTORRGTKGKPGIYRFVAPGERPSPGMSVLCVYDAG 562  
Db 1469 SLDPFTTITITLPQDAVSRTORRGTKGKPGIYRFVAPGERPSPGMSVLCVYDAG 1528  
QY 563 CAWYELTPAETTVRLRAYNTPGLPVCQDHLBFWEVFTGLTHIDAHFLSQTKQSGENLP 622  
Db 1529 AAWYELTPAETTVRLRAYNTPGLPVCQDHLBFWEVFTGLTHIDAHFLSQTKQSGENFA 1588  
QY 623 YLVAYOATVCARQAQPPSPMDQWKKLIRLKPTLHGPTLLYRLGAVQNEITLTHPVTKY 682  
Db 1589 YLTAYOATVCARAKAPPPSPMDVWKKLIRLKPTLHGPTLLYRLGAVTNEVTLTHPVTKY 1648  
QY 683 IMTMSADLEWVTSACSGKPAIIPDREVLRYRFDEMECSQHLPIYEQGMALAEQFKQA 742  
Db 1649 IATCMQADLEWTS-----VAAYCLATGCIS-----IIGRLHLNDRVV 1697  
QY 802 VTPDKELYEAFDEMECSKAALTEEGORMAEMLSKIOGLL----- 844  
Db 1698 VAPDKELYEAFDEMECSKAALTEEGORMAEMLSKIOGLLQQAQROAQDIQPAIQSS 1757  
QY 845 ----- 844  
Db 1758 WPKLEQFWAKHWNFIISGTYLAGLSTLPGNPVAVSMMAFSAALTSPLTSTTILLINMG 1817  
QY 845 ----- 844

```
Db 1818 GWLASQIAPPAGATGTFVVGSLGNAVGSIGLGKILVDVLGAGYAGIGAGISGALVAFKINSGEK 1877
Qy 845 -----GILRRHYGPGEGAGVQWMNRLIAFASRGNHVSPTH 878
Db 1878 PTVEDVVNLLPAILSPGALVGVICAAILRRHHYQGEGAGVQWMNRLIAFASRGNHVAPTH 1937
Qy 879 YV--PSRRRPAQALP----- 892
Db 1938 YVVEDASORVTQVLSLTTITSLRLRLHAWITEDCPVPCSGSLQDIWDWCSILTDFKN 1997
Qy 893 -----VWA----- 895
Db 1998 WLSKLLPKMPGIPFISCGQYKGVAGTGMVTRCPCGANISGHVRKMTKGTGPKTCL 2057
Qy 896 -----RPDYNPLVET--WK----- 908
Db 2058 NLWQGTPIPCYTEGCPVKP---PPNYKTAIMRVAASEYVEVTHQGSFSVTLGTSNLT 2114
Qy 909 -----KPDYE-- 913
Db 2115 KVPQVPAPBFFSWVDGVQIHRFAPVPGPPFRDEVTFTVGLNSFVVGSQLPCDPEDPTEV 2174
Qy 914 -----PPVHGRSS----- 922
Db 2175 LASMLTDPSHITAEAAARRLARGSPSQASSASQLSAPSLKATCTTHKTAAYDCDMVDAN 2234
Qy 923 -----RRFAQALPVWARPDPYN 938
Db 2235 LFMGDDVTRIESDKVIVLSDLSMTVEDDREPSVSEVYLKERRKPPALPPWARPDPYN 2294
Qy 939 PPLVETWKPDYEPVPHVGHGKTKRNTNRPODVKFGGQIVGRRGPPIPKARRPEGRTW 998
Db 2295 PVLITWKRCYEPETVLG-----CALPPTPQ-----TPVPPRRRRRAKVL 2335
Qy 999 AQCYPW-----PLYGNKDR-RSTGKSWG 1021
Db 2336 TDONVEGVLEMAKDVLSPLQDNNDSGHSTGADTG 2370

RESULT 12
S68016
ATPase/RNA helicase - hepatitis C virus (fragment)
C;Species: hepatitis C virus
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 09-Jul-2004
C;Accession: S68016
R;Jin, L.; Peterson, D.L.
Arch. Biochem. Biophys. 323, 47-53, 1995
A;Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA
A;Reference number: S68016; MUID:96019946; PMID:7487072
A;Accession: S68016
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-386 <JIN>
A;Cross-references: UNIPROT:Q04045
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polyprotein
F;24-31/Region: nucleotide-binding motif A (P-loop)
F;86-91/Region: nucleotide-binding motif B
F;90-93/Region: DEXH motif

Query Match 33.5%; Score 1980.5; DB 2; Length 386;
Best Local Similarity 92.6%; Pred. No. 2.4e-116;
Matches 376; Conservative 4; Mismatches 5; Indels 21; Gaps 2;

Qy 246 PVFTDN-SSPPVVDQSFOVAHLHAPTSGSKTKVPAAYAAQYKVLNPSVAATLFGA 304
Db 1 PVFTDNSSPPVVDQSFOVAHLHAPTSGSKTKVPAAYAAQYKVLNPSVAATLFGA 60
Qy 305 YMSKAHGIDPNIRGTITTTGSPITTYSTYKFLADGCGSGAYDIIICDCHSTDATSI 364
Db 61 YMSKAHGV-----YKFLADGCGSGAYDIIICDCHSTDATSI 100
Qy 365 LGICTVLDDQABTAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEVIK 424
```

```
Db 101 LGICTVLDDQABTAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEAIK 160
Qy 425 GGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTG 484
Db 161 GGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGFSG 220
Qy 485 DFDVSIDCNTCVTQTVDFSLDPTFTTITLPODAVSRTOGRTGKGGIYRFVAPGE 544
Db 221 DFDVSIDCNTCVTQTVDFSLDPTFTTITLPODAVSRTOGRTGKGGIYRFVAPGE 280
Qy 545 RPSGMPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPCLPVCQDHLFEWEGVFTGLT 604
Db 281 RPSGMPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPCLPVCQDHLFEWEGVFTGLT 340
Qy 605 HIDAHLFSLQTKQSGENLPYLVAQVCAQAQAPPPSWDQMKCL 650
Db 341 HIDAHLFSLQTKQSGENFPYLVAQVCAQAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 386

RESULT 13
PC2219
polypeptide - hepatitis C virus (type 5a) (fragments)
N;Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote
C;Species: hepatitis C virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C;Accession: PC2219
R;Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994
A;Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the h
A;Reference number: PC2219; MUID:94338342; PMID:7520237
A;Accession: PC2219
A;Molecule type: mRNA
A;Residues: 1-876 <STU>
A;Cross-references: UNIPROT:O81242; GB:L295577; GB:L295578; GB:L295579
A;Experimental source: serum
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: glycoprotein
F;1-191/Product: core #status predicted <COE>
F;68-78/Region: variable
F;192-247/Product: E1 (carboxyl end) #status predicted <ERE>
F;248-411/Product: E2/NS1 (amino end) #status predicted <ENR>
F;248-338/Region: E2
F;339-411/Region: NS1 (amino end)
F;412-783/Product: NS3 #status predicted <NSR>
F;784-837/Product: NS4A #status predicted <NSA>
F;838-876/Product: NS4B #status predicted <NSB>
F;281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 33.3%; Score 1966.5; DB 2; Length 876;
Best Local Similarity 52.5%; Pred. No. 6.2e-115;
Matches 426; Conservative 66; Mismatches 148; Indels 171; Gaps 18;

Qy 69 SRKGGPKDERHVGDLGNTADKQ-CVADVSIEDSVISLGDHCIIIGTLVWHEKADDL 127
Db 103 SRPNMGNDPRKRSNRNKGKVIDTLTCGFAD-----LMGYPLVGGPI----- 144
Qy 128 GKGGNEBTKTGNAAGSLACVIGIAQLN-----SGCNCISY-----PGH--- 168
Db 145 -----GGVARALAHGVRLVDGNYATGNLPCCSFSIFILALLSLCTVTPASAPT 193
Qy 169 -----ITGHRMAWK-----GSAARTTSFVSLFPAKQONETHV 203
Db 194 TALLVAQLLRIPQVVDIIAGSH--WGVLFAAAYASVANNWTKVVLVFLFAGVDATTQI 251
Qy 204 TGGAAATTSGLTSLFPGASQNIQLI-----VDIPV-----ENLET----- 241
Db 252 SGGSAQTITGIASFITRGAQQLQINTNGSMHINRTALNCNDSLOTGTFIAGFYFHKF 311
Qy 242 -----TMRSPVFTD-----NSPPVVDQSFOVAHLHAPTSGSKTKVPAAY 282
Db 312 NSSCCPDMASCRALAFDQGWGTISYANISGPDSDPKYCHYPPRCG-----VVPAGE 366
```



```
QY 283 AAQGYKVLNPSVAATLFGAYMSKAHGIDPNIRTVRT-----ITTGSPIT 330
Db 367 VGPVVCFTSPVVGTTD-----SKGH---PTYNMGSNVTFDFLMNTRPPIITGASIT 418
QY 331 YSTYKFLADGGCGSGAYDIIICDECHSDATSIILGIGTVLDQAEATAGARLVVLATATPP 390
Db 419 YSTYKFLADGGCGSGAYDIIICDECHSQDATIILGIGTVLDQAEATAGARLVVLATATPP 478
QY 391 GSVTPHPNIEVALSTTGEIPYGAIPLEVIKGRHLIFCHSKKKKDELAALVALGI 450
Db 479 GSVTTPHPNIEVALPQEGEVPPYGRAIPLAFIKGRHLIFCHSKKKKDELAQLTSLGV 538
QY 451 NAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVIDCNTCTVTVDFSLDPTFTI 510
Db 539 NAVAYRGLDVAIVPTAGDVVVVCTDALMTGFTGFDSVIDCNSAVTQTVDFSLDPTFTI 598
QY 511 ETITLPQDAVSRQRRGRGKPGIYRFVAPGERPSGMFDDSVLCECYDAGCAWYELTP 570
Db 599 ETITVPQDAVSRQRRGRGKPGIYRYVSAGERPSDMFDSVVLCECYDAGCAWYDLTP 658
QY 571 AETTIVRLRAYMNTPGLPVQCDHLEFWEVFTGLTHIDAHFLSQTKSGENLPYLVAIQAT 630
Db 659 AETTIVRLRAYINTPGLPVQCDHLEFWEVFTGLTHIDAHFLSQTKSGENFPYLVAIQAT 718
QY 631 VCARAQAPPSWDMMKCLIRLKPITLHGPTLLYRLGAVQNEITLTHPVTKYIMTCSAD 690
Db 719 VCVRAKAPPSWDTMWKMLRLKPTLTGPTLLYRLGVPQNEITLTHPIKTYIMACSAD 778
QY 691 LEVWISA-----CSGKPAIIPDREVLYREFDEMECS 722
Db 779 LEVITSTWVLVGGVAAALAAAYCLTVGSVAIVGRILISGKPAIIPDREALYQQFDEMECS 838
QY 723 QHLPTYEQGMMLABQPKQKALG-LSRGGKPA 752
Db 839 ASLPYMDETRAIAQGFKEKVLGPISTTGQA 869

RESULT 14
PS0326
polyprotein - hepatitis C virus (isolate Fla) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: PS0326
R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.
Gene 105, 167-172, 1991
A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype
A:Reference number: PS0326; MUID:92039028; PMID:1718820
A:Accession: PS0326
A:Molecule type: genomic RNA
A:Residues: 1-492 <LiJ>
A:Cross-references: UNIPROT:Q9IFES; UNIPROT:Q36579; UNIPROT:Q36610; UNIPROT:Q03463; UNIPROT:M60220
A:Note: this sequence corresponds to nonstructural protein NS3 region
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 25.1%; Score 1483.5; DB 2; Length 492;
Best Local Similarity 62.4%; Pred. No. 4.3e-85;
Matches 306; Conservative 22; Mismatches 63; Indels 99; Gaps 8;

QY 488 SVIDCNCVTQTVDFSLDPTFTIETITLPQDAVSRQRRGRGKPGIYRFVAPGERPS 547
Db 1 SVIDCNCVTQTVDFSLDPTFTIETITLPQDAVSRQRRGRGKPGIYRFVAPGERPS 60
QY 548 GMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCDHLEFWEVFTGLTHID 607
Db 61 GMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCDHLEFWEVFTGLTHID 120
QY 608 AHFLSQTKSGENLPYLVAIQATVCARAQAPPSWDMMKCLIRLKPITLHGPTLLYRLG 667
Db 121 AHFLSQTKSGENLPYLVAIQATVCARAQAPPSWDMMKCLIRLKPITLHGPTLLYRLG 180
```

```
QY 668 AVQNEITLTHPVTKYIMTCSADLEVVTS-----ACS 699
Db 181 AVQNEITLTHPVTKYIMTCSADLEVVTSVWLVGVLAAALAAAYCLSTGCVVIVGRVLS 240
QY 700 GKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLABQPKQKALG-----SRGGKPAIYP 755
Db 241 GKPAIIPDREVLYQSFDEMECSQHLPIYEQGMMLABQPKQKALGLLOQTASRAE-AITP 299
QY 756 DREVLYQQYD-----EMEECSQAAPYISQAQVIAHQFKEKVLGLINDQVVVTP---DKE 807
Db 300 AVQTNQWRQLETFWAKHMNMFISGIQYLAGLSTLPG--NPATIASLMAFTAAVTSPLTTSOT 357
QY 808 ILYE-----AFDEMECASKAALIEEGORMAEMLKSKIQGLLG-----845
Db 358 LLFNILGHWAAQAALAPGNATAFVAGLAGAIGAIVGLGKVLVDILAGYGVAGALVAF 417
QY 846 -----ILRRHVGPGEGAVQNMNRLIAPASRG 871
Db 418 KIMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVQNMNRLIAPASRG 477
QY 872 NHVSPTHYVP 881
Db 478 NHVSPTHYVP 487

RESULT 15
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: JQ1366
R:Krensdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implications
A:Reference number: JQ1366; MUID:92013977; PMID:1655961
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>
A:Cross-references: UNIPROT:Q9PX22
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #state

Query Match 23.5%; Score 1386.5; DB 2; Length 716;
Best Local Similarity 46.6%; Pred. No. 8.8e-79;
Matches 336; Conservative 51; Mismatches 115; Indels 219; Gaps 27;

QY 200 ETHVTGGAARTTSGLTSLFSPCASQNIQLI-----VDFIPV---ENLETTMRSPVFT 249
Db 51 EIVTGTGSTARITQGLVSLFSRGAQDIQLINTNGSMHINRTALNCNESLDTGWAGLGF- 109
QY 250 DNSSPVPVPSQFVAHLHAPTSGSKSTKV-----PAAYAAQGYKVLNPSVAATLFGA- 304
Db 110 -----YYHKFNSSGCPERMASCRPLADFDQGM-----GPISYANGTGE 148
QY 305 -----YMSKAHGIDPNIRTVRTIT-----TGSPITYSTYVKFLA---DGCSCGGAY 348
Db 149 HRPYCHWHPKPGIYV-----AQTVCGPVYCFPTSPVVVGITNKLGAPTYNMGCD--T 201
QY 349 DIII-----CDECHSDATSIILGIGTVLDQAEATAGARLVVLATATPP----- 390
Db 202 DVFVLANTRPPLGNWFGCTWNSSGFTKVCG-----APPCVIG 240
QY 391 -GSVTV-----PHPNIEVALSTTGEIIPF-----YGAIAIPLVLIK-- 424
Db 241 AGNNTLYCPTDCFRKHP---EATYRCGSGPMITPRLCVGYPVRLMHPYCTVNYTLFKVR 297
QY 425 ---GG---RHLIFCH--SKKCC---DELAACLVALGINAVYVRGLDVSIVPTSGDVVV 473
Db 298 MYVGVGVEHRLQVACNWTRGERCNLDRDRSELSPLLSITQW-----QVLPCS-----FT 347
QY 474 ATDALMTGYTGDPSVID-----C-NTCV 496
```



Db	348	TLPALTTGLHLHQNIVDVQYLYGVGSSIVSWAIKWEYVILLFLLADARVCSCLMNTCV	407
Qy	497	TQTVDFSLDPTFTIETITLQDAVSRTRGRGKPGIYRFVAPGERPSGMFDSVLC	556
Db	408	TQTVDFSLDPTFTIETITLQDAVSRTRGRGKPGIYRFVAPGERPSGMFDSVLC	467
Qy	557	ECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSOTKQ	616
Db	468	ECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSOTKQ	527
Qy	617	SGENLPYLVAQATVCARAQAPPSWDMKCLIRLAKPTLHGPTPLLYRLGAVQNEITLT	676
Db	528	SGENLPYLVAQATVCARAQAPPSWDMKCLIRLAKPTLHGPTPLLYRLGAVQNEVILT	587
Qy	677	HPVTKYIMTMSADLEVVVISA-----CSGKPAIIPDR	708
Db	588	HPITKYIMTMSADLEVVVTSWTLVGGVLAALAAAYCLSTGCVIVGRVILSGKPAIIPDR	647
Qy	709	EVLRYREFDEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQOY	764
Db	648	EVLRYREFDEMECSQHLPIYIEQGMMLAEQFKQKALGLLQTRSRQAE-VITPAVQTNQRL	706
Qy	765	D 765	
Db	707	E 707	

Search completed: November 7, 2005, 20:11:43  
Job time : 48.7031 secs

This page is blank.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 140.844 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-6  
Perfect score: 5912  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRSTGKSGKPGYPWP 1099

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2909.5	49.2	2436	Q81756	Q81756 hepatitis c
2	2909.5	49.2	3011	1 POLG HCV1	P26664 h genome po
3	2905.5	49.1	3011	2 Q3IFES	Q3IFES hepatitis c
4	2875	48.6	3011	2 Q36F79	Q36F79 hepatitis c
5	2868	48.5	3011	2 Q36608	Q36608 hepatitis c
6	2868	48.5	3011	2 Q36610	Q36610 hepatitis c
7	2868	48.5	3015	2 Q9PMW9	Q9PMW9 hepatitis c
8	2868	48.5	3015	2 Q9PMW5	Q9PMW5 hepatitis c
9	2866	48.5	2908	2 Q6IX04	Q6IX04 hepatitis c
10	2861	48.4	3011	2 Q36609	Q36609 hepatitis c
11	2851	48.2	3011	2 Q36ELS8	Q36ELS8 hepatitis c
12	2850	48.2	3011	2 Q3DIT6	Q3DIT6 hepatitis c
13	2835	48.0	3011	2 Q3DIT6	Q3DIT6 hepatitis c
14	2821	47.7	3011	1 POLG HCVH	P27958 h genome po
15	2783.5	47.1	3010	2 Q9J3H0	Q9J3H0 hepatitis c
16	2777.5	47.0	3010	2 Q9JTE8	Q9JTE8 hepatitis c
17	2775	46.9	3010	2 Q9J3G4	Q9J3G4 hepatitis c
18	2770	46.9	3013	2 Q9QNC0	Q9QNC0 hepatitis c
19	2768.5	46.8	3010	1 POLG HCVJT	P00269 h genome po
20	2766	46.8	3014	2 Q9DTE0	Q9DTE0 hepatitis c
21	2764.5	46.8	3010	2 Q9J3G8	Q9J3G8 hepatitis c
22	2764	46.8	3010	2 Q9J3G2	Q9J3G2 hepatitis c
23	2763.5	46.7	3010	2 Q9QIY2	Q9QIY2 hepatitis c
24	2763	46.7	3010	2 Q9QIY5	Q9QIY5 hepatitis c
25	2763	46.7	3010	2 Q9QIY6	Q9QIY6 hepatitis c
26	2762.5	46.7	3010	2 Q9QIX2	Q9QIX2 hepatitis c
27	2762.5	46.7	3010	2 Q9QIY1	Q9QIY1 hepatitis c
28	2761	46.7	3010	2 Q9DTE6	Q9DTE6 hepatitis c
29	2759	46.7	3010	2 Q9DTE9	Q9DTE9 hepatitis c
30	2759	46.7	3010	2 Q9J3H8	Q9J3H8 hepatitis c
31	2758	46.7	3010	2 Q81757	Q81757 hepatitis c

ALIGNMENTS				
RESULT 1				
ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M32084; AAA45677.1; -			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P27958; 1A1V.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0019038; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006350; P:transcription; IEA.			
DR	GO; GO:0019079; P:viral genome replication; IEA.			
DR	GO; GO:0019087; P:viral transmembrane; IEA.			
DR	InterPro; IPR000345; CytC_heme_BS.			
DR	InterPro; IPR004410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002531; HCV NS1.			
DR	InterPro; IPR000745; HCV NS4a.			
DR	InterPro; IPR002868; HCV NS4b.			
DR	InterPro; IPR002166; HCV NS5a.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR009003; Pept Ser Cys.			
DR	InterPro; IPR002518; Pept U39_HCV NS2.			
DR	InterPro; IPR007095; RNA_pol_PS.			
DR	InterPro; IPR007094; RNA_pol_PSVir.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	Pfam; PF01538; HCV NS2; 1.			
DR	Pfam; PF02907; HCV NS3; 1.			
DR	Pfam; PF01006; HCV NS4a; 1.			
DR	Pfam; PF01001; HCV NS4b; 1.			
DR	Pfam; PF01506; HCV NS5a; 1.			



DR PDB; 1HEI; X-ray; A/B=1206-1656.  
 DR PDB; 1ONB; NMR; A=1349-1507.  
 DR MEROPS; S29.001; -  
 DR MEROPS; U39.001; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 DR InterPro; IPR004109; Peptidase\_S23.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR 3D-structure; ATP-binding; Coat protein; Core protein;  
 DR Envelope protein; Glycoprotein; Helicase; Hydrolase;  
 DR Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 DR Serine protease; Transferase; Transmembrane.  
 DR INIT\_MST 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 283  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 FT STRAND 1224 1226  
 PDB; 1HEI; X-ray; A/B=1206-1656.  
 PDB; 1ONB; NMR; A=1349-1507.  
 MEROPS; S29.001; -  
 MEROPS; U39.001; -  
 InterPro; IPR001410; DEAD.  
 InterPro; IPR002522; HCV\_capsid.  
 InterPro; IPR002521; HCV\_core.  
 InterPro; IPR002519; HCV\_env.  
 InterPro; IPR002531; HCV\_NS1.  
 InterPro; IPR000745; HCV\_NS4a.  
 InterPro; IPR001490; HCV\_NS4b.  
 InterPro; IPR002868; HCV\_NS5a.  
 InterPro; IPR002166; HCV\_RdRP.  
 InterPro; IPR001650; Helicase\_C.  
 InterPro; IPR009003; Pept\_Ser\_Cys.  
 InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 InterPro; IPR004109; Peptidase\_S23.  
 InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 InterPro; IPR007094; RNA\_pol\_PSVir.  
 Pfam; PF01543; HCV\_capsid; 1.  
 Pfam; PF01542; HCV\_core; 1.  
 Pfam; PF01539; HCV\_env; 1.  
 Pfam; PF01560; HCV\_NS1; 1.  
 Pfam; PF01538; HCV\_NS2; 1.  
 Pfam; PF02907; HCV\_NS3; 1.  
 Pfam; PF01006; HCV\_NS4a; 1.  
 Pfam; PF01001; HCV\_NS4b; 1.  
 Pfam; PF01506; HCV\_NS5a; 1.  
 Pfam; PF00271; Helicase\_C; 1.  
 Pfam; PF00998; Viral\_RdRP; 1.  
 SMART; SM00487; DEXDC; 1.  
 3D-structure; ATP-binding; Coat protein; Core protein;  
 Envelope protein; Glycoprotein; Helicase; Hydrolase;  
 Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 Serine protease; Transferase; Transmembrane.  
 INIT\_MST 1  
 CHAIN 1 115  
 CHAIN 116 191  
 CHAIN 192 283  
 CHAIN 384 729  
 CHAIN 730 1006  
 CHAIN 1007 1615  
 CHAIN 1616 1862  
 CHAIN 1863 2013  
 CHAIN 2014 3011  
 TRANSMEM 347 369  
 ACT\_SITE 1083 1083  
 ACT\_SITE 1107 1107  
 ACT\_SITE 1165 1165  
 NP\_BIND 1230 1237  
 SITE 1316 1319  
 CARBOHYD 196 196  
 CARBOHYD 209 209  
 CARBOHYD 234 234  
 CARBOHYD 305 305  
 CARBOHYD 417 417  
 CARBOHYD 423 423  
 CARBOHYD 430 430  
 CARBOHYD 448 448  
 CARBOHYD 476 476  
 CARBOHYD 532 532  
 CARBOHYD 540 540  
 CARBOHYD 556 556  
 CARBOHYD 576 576  
 CARBOHYD 623 623  
 CARBOHYD 645 645  
 CARBOHYD 2041 2041  
 CARBOHYD 2077 2077  
 CARBOHYD 2240 2240  
 CARBOHYD 2364 2364  
 CARBOHYD 2789 2789  
 STRAND 1224 1226

FT TURN 1234  
 FT TURN 1236  
 FT HELIX 1238  
 FT TURN 1246  
 FT TURN 1247  
 FT TURN 1248  
 FT STRAND 1251  
 FT STRAND 1255  
 FT STRAND 1291  
 FT STRAND 1295  
 FT HELIX 1301  
 FT TURN 1303  
 FT STRAND 1312  
 FT STRAND 1315  
 FT TURN 1316  
 FT TURN 1319  
 FT HELIX 1323  
 FT TURN 1335  
 FT TURN 1336  
 FT HELIX 1337  
 FT TURN 1340  
 FT STRAND 1343  
 FT TURN 1346  
 FT TURN 1352  
 FT STRAND 1353  
 FT STRAND 1363  
 FT STRAND 1366  
 FT STRAND 1373  
 FT TURN 1377  
 FT STRAND 1380  
 FT HELIX 1382  
 FT STRAND 1385  
 FT STRAND 1390  
 FT HELIX 1393  
 FT TURN 1397  
 FT TURN 1408  
 FT TURN 1411  
 FT STRAND 1414  
 FT TURN 1417  
 FT TURN 1430  
 FT STRAND 1431  
 FT STRAND 1433  
 FT TURN 1436  
 FT TURN 1438  
 FT TURN 1439  
 FT TURN 1442  
 FT STRAND 1444  
 FT STRAND 1450  
 FT STRAND 1453  
 FT STRAND 1456  
 FT STRAND 1471  
 FT STRAND 1478  
 FT HELIX 1481  
 FT TURN 1488  
 FT TURN 1490  
 FT STRAND 1498  
 FT STRAND 1501  
 FT HELIX 1514  
 FT TURN 1526  
 FT TURN 1527  
 FT HELIX 1532  
 FT TURN 1544  
 FT HELIX 1546  
 FT TURN 1547  
 FT HELIX 1555  
 FT TURN 1563  
 FT TURN 1564  
 FT TURN 1564  
 FT HELIX 1570  
 FT TURN 1578  
 FT TURN 1580  
 FT HELIX 1584  
 FT TURN 1597  
 FT TURN 1598  
 FT HELIX 1598  
 FT TURN 1606  
 FT TURN 1611  
 FT TURN 1618  
 FT STRAND 1623  
 FT STRAND 1627  
 FT STRAND 1635  
 FT HELIX 1636  
 FT HELIX 1640  
 FT HELIX 1652  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;  
 Query Match 49.2%; Score 2909.5; DB 1; Length 3011;  
 Best Local Similarity 53.3%; Pred.No. 1.1e-165;  
 Matches 622; Conservative 33; Mismatches 92; Indels 419; Gaps 16;  
 QY 224 SQNIQLIVDFIPVENLETMTSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAVA 283  
 Db 1186 TRGVAKAVDFIPVENLETMTSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAVA 1245  
 QY 284 AQGYKVLVLPNSVAATLGFQAYMSKAHGIDNPRTGVTITTTGSPITTYSTYKFLADGGC 343  
 Db 1246 AQGYKVLVLPNSVAATLGFQAYMSKAHGIDNPRTGVTITTTGSPITTYSTYKFLADGGC 1305  
 QY 344 SGGAYDIIICDECHSTDATSIIGICTVLDQAEATAGARLVVLATATPPGSVTVPHNPINIEV 403  
 Db 1306 SGGAYDIIICDECHSTDATSIIGICTVLDQAEATAGARLVVLATATPPGSVTVPHNPINIEV 1365  
 QY 404 ALSTTGEIPFFYGYKAIPLEVIKGRHLIFCHSKKKCDELAALKLVALGINAVAYYRGLDVSV 463  
 Db 1366 ALSTTGEIPFFYGYKAIPLEVIKGRHLIFCHSKKKCDELAALKLVALGINAVAYYRGLDVSV 1425

```
Qy 464 IPTSGDVVVVATDALMTGYTGDFDSVIDDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRT 523
Db 1426 IPTSGDVVVVATDALMTGYTGDFDSVIDDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRT 1485
Qy 524 ORRGTRGRGPGIYRFVAPGRPSGMEDSSVLCYDAGCAWYELTFAETTVRLRAYMNT 583
Db 1486 ORRGTRGRGPGIYRFVAPGRPSGMEDSSVLCYDAGCAWYELTFAETTVRLRAYMNT 1545
Qy 584 PGLPVCODHLEFWEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSWD 643
Db 1546 PGLPVCODHLEFWEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSWD 1605
Qy 644 QMVKCLIRLKLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS ----- 696
Db 1606 QMVKCLIRLKLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSVWLVG 1665
Qy 697 -----ACSGKPAIIPDREVLVREDEMEECQHLPLYTEQGMMLA 735
Db 1666 VLAALAAAYCLSTGCVTVGRVLSGKPAIIPDREVLVREDEMEECQHLPLYTEQGMMLA 1725
Qy 736 EQFKQKALGL-----SRGKPAIIPDKEVLVQOYD-----EMEESQAAPVIEQAQVIAHQ 786
Db 1726 EQFKQKALGLQATASROAE-VIAPAVQTNQKLETFWAKHWNFIISQIYVLGSLTLPG- 1783
Qy 787 FKSKVLGLINDQVVVTP-----DKEILYE-----AFDEMEECASKAALI 826
Db 1784 -NPAIASLMAFTAATAVTSPLTTSQTLLFNILGWMVAQAAPGAATAFVGAGLAGAIGSV 1842
Qy 827 EEQRMWAEMLKSKTOGLLG-----IL 847
Db 1843 GLGKVLIDILAGVAGVAGALVAFKIMSGEVFSTEDLVNLLPALSPGALVWGVCAAIL 1902
Qy 848 RRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPS----- 882
Db 1903 RRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPSDAAARVAILSSLTVTQLLRHLQ 1962
Qy 883 ----- 882
Db 1963 WISSECTTPCSGWLRIWDWICEVLSDFKTLKAKLMPQLPGIPFVSCQRYGKGVWVD 2022
Qy 883 ----- 882
Db 2023 GIMHTRCHGAEITHVKNGTMRIVGPRTCRNMMSTGFPINAVYTGCTELPAPNVTAL 2082
Qy 883 -----RSRFEA----- 888
Db 2083 WRVSAEYVEIRQVGDHYVTGTTDNLKCPCQVPSEFFTELDGVRLHRAFPCKPLLR 2142
Qy 889 -----QALPVWARPDY-----NPP----- 902
Db 2143 EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAAGRLARGSPFSVASS 2202
Qy 903 -----LVET---WKK-----PYEPPVH 918
Db 2203 ASQLSAPSLKATCANHDSPEALIEANLLWRQEMGNNITRVSEKNKWLIDSFDELVAE 2262
Qy 919 G-----RSSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHVGRKTKRNTNR 966
Db 2263 EDREISVPAEILKSRFAQALPVWARPDPNPLVETWKKPDYEPVPHV----- 2312
Qy 967 RPODVKPPGGQIVGRGPPPKARR 992
Db 2313 -----GCPLPPKSPVPPPPRK 2329
```

## RESULT 3

```
Q9IF5 PRELIMINARY; PRT; 3011 AA.
AC Q9IF5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.

Query Match 49.1%; Score 2905.5; DB 2; Length 3011;
Best Local Similarity 53.3%; Pred. No. 1.9e-165;
Matches 621; Conservative 34; Mismatches 92; Indels 419; Gaps 16;

Qy 224 SQNIQLIVDFIPVENLETTMRSPVTDNSSPPVPSFOVAHLHAPTSGSKTKVPAAYA 283
Db 1186 TRGKAVAKVDFIPVENLETTMRSPVTDNSSPPVPSFOVAHLHAPTSGSKTKVPAAYA 1245
```

```
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
DR EMBL; AF271632; AAF81759.1; -.
DR PIR; A44150; A44150.
DR PIR; Q0804; Q0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSP; QBUYS1; ICWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR00345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_env.
DR InterPro; IPR002519; HCV NS1.
DR InterPro; IPR002531; HCV NS4.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept Ser Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;
```

```
Qy 284 AQGYKVLVLPNSVAATLGFAGYMSKAHGIDPNIRTVGRTTITGSPITYSTYTGKFLADGGC 343
Db 1246 AQGYKVLVLPNSVAATLGFAGYMSKAHGIDPNIRTVGRTTITGSPITYSTYTGKFLADGGC 1305
Qy 344 SGGAYDIIICDECHSDTATILGIGTVLDOAETAGARLVLAATATPGSVTVPHNLEEV 403
Db 1306 SGGAYDIIICDECHSDTATILGIGTVLDOAETAGARLVLAATATPGSVTVPHNLEEV 1365
Qy 404 ALSTTGIPYGYKAIPLEVIKGRHLLIFCHSKKKDELAALVALGINAVAYRGLDVSV 463
Db 1366 ALSTTGIPYGYKAIPLEVIKGRHLLIFCHSKKKDELAALVALGINAVAYRGLDVSV 1425
Qy 464 IPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVTQTVDFSLDPTFTTITLPODAVSRT 523
Db 1426 IPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVTQTVDFSLDPTFTTITLPODAVSRT 1485
Qy 524 QRRGRTGRGPGIYRFVAPGERSGMDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT 583
Db 1486 QRRGRTGRGPGIYRFVAPGERSGMDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT 1545
Qy 584 PGLPVCQDHLFEFMEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVTCARQAAPPSPSWD 643
Db 1546 PGLPVCQDHLFEFMEGVFTGLTHIDAHFLSQTQKSGENLPYLVAQVTCARQAAPPSPSWD 1605
Qy 644 QMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS 696
Db 1606 QMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS 1665
Qy 697 -----ACSGKPAIPDRVLYREDEMECSQHLPIYIEQGMWLA 735
Db 1666 VLAALAAAYCLSTGCWIVGRVVLVSGKPAIPDRVLYREDEMECSQHLPIYIEQGMWLA 1725
Qy 736 EGFQKQKALGL-----SRGKPAIPDPKEVLYQOYD-----EMECSQAAPYIEQAQVIAHQ 786
Db 1726 EGFQKQKALGLLQTSARQAE-VIAFVQTNWQKLETFWAKHWNVTSIGIYLAGLSTLPG- 1783
Qy 787 FKEKVLGLINDQVVVTP-----DKELIYE-----AFDEMECSKAKALI 826
Db 1784 -NPAISLMAFTAATAVTSPLTTSQTLLFNILGWVAAQLAAPGAATAFVGLAGLAGAAGSV 1842
Qy 827 EGGQMAEMLSKTKQGLLG-----IL 847
Db 1843 GLGKVLIDILAGYGAGVAGALVAFKIMSGEVSPSTEDLVNLLPAILSPGALVGVVCAAIL 1902
Qy 848 RRHVGPGEVAGVQWNNRLIAFASRGHNHVSPTHYVPS----- 882
Db 1903 RRHVGPGEVAGVQWNNRLIAFASRGHNHVSPTHYVPSDDAAARVAILSSLTVTQLRLHQ 1962
Qy 883 ----- 882
Db 1963 WISSECTTPCSGSLRDIWDWICEVLSDFKTWLKAQLMPQLPGIPFVSCQRYKGVWRGD 2022
Qy 883 ----- 882
Db 2023 GIMHTRCHGAEITGHVKNGTMRIVGPRTCRNMMWSTGTFPINAYTTGPTCTPLPAPNYTFAL 2082
Qy 883 -----RSRREA----- 888
Db 2083 WRVSAEEVEITRQVGDHFHYVTGMTDNLKCPQVPSPEFFTELDGVLRLHRFAPCKPLLR 2142
Qy 889 -----QALPVWARPDY-----NPP----- 902
Db 2143 EEVSFRVGLHEYPVCSQLPCEPEPDVAVLTSMLTDPDSHITAEAGRRRLARGSPPSVASSS 2202
Qy 903 -----LVET-----WKK-----PDVEPPVH 918
Db 2203 ASQLSAPSLKATCTANHDSPDALTEANLLRQEMGNITRVESKNKVVILDSFDPLVAE 2262
Qy 919 G-----RSSRRFAQALPVWARPDPNPLVETWKKPDPYPPVHGRKTKRNTNR 966
Db 2263 EDEREISVPAELLRSSRRFAQALPVWARPDPNPLVETWKKPDPYPPVH----- 2312
Qy 967 RPQDVKFPGGGIVGRRGPPIPKARR 992
```

```
Db 2313 -----GCPLPPPKSPVPPPRK 2329
RESULT 4
O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RT "Transmission of hepatitis C by intrahepatic inoculation with
transcribed RNA.";
RL Science 277:570-574(1997).
DR EMBL; AF009606; AAB6324.1; -.
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR PDB; 1N1L; X-ray; A/B=1017-1214.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR011410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
```



KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327184 MW; E2E0EB809C63C1B9 CRC64;

Query Match 48.6%; Score 2875; DB 2; Length 3011;  
Best local similarity 51.8%; Pred. No. 1.3e-163;  
Matches 617; Conservative 41; Mismatches 104; Indels 430; Gaps 17;

QY 204 TGGAAATTTSGLTSPG-ASQNIQLIYDFIPVENLETTMTSPVFTDSSPPVPOSFQ 262  
DB 1165 SGGLPLCPAGHAGVGLFRAAVCTRGVAKAVDFIPVENLETTMTSPVFTDSSPPVPOSFQ 1224

QY 263 VAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGIDPNIRTVRT 322  
DB 1225 VAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLFGAYMSKAHGVDPNIRTVRT 1284

QY 323 ITTGSPIITYSTYKFLADGCGSGGAYDIIICDECHSTDATSIIGTIGTVLDOAETAGARLV 382  
DB 1285 ITTGSPIITYSTYKFLADGCGSGGAYDIIICDECHSTDATSIIGTIGTVLDOAETAGARLV 1344

QY 383 VLATATPPGVTVPHPNIEBALSTTGEIPFYKAIPLEVIKGRHLIFCHSKKKDELA 442  
DB 1345 VLATATPPGVTVSHPNIEBALSTTGEIPFYKAIPLEVIKGRHLIFCHSKKKDELA 1404

QY 443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDVSDVIDCNTCTVTQVDF 502  
DB 1405 AKLVALGINAVAYRGLDVSIVPTSGDVVVVSTDALMTGYTGDVSDVIDCNTCTVTQVDF 1464

QY 503 SLDPTFTIETITLPQDAVSTORRGTRGKPGIYRFVAPGERPSGMFDSVLCVCYDAG 562  
DB 1465 SLDPTFTIETITLPQDAVSTORRGTRGKPGIYRFVAPGERPSGMFDSVLCVCYDAG 1524

QY 563 CAMELTTPAETTVRLRAYMTPGLPVCDHLEFWGVFTGLTHIDAHFLSQTOSGENLP 622  
DB 1525 CAMELTTPAETTVRLRAYMTPGLPVCDHLEFWGVFTGLTHIDAHFLSQTOSGENFP 1584

QY 623 YLVAYQATVCARAQAQPPSPDQMWKCLIRLKTPLHGTPTLLYRLGAVQNEITITHPVTKY 682  
DB 1585 YLVAYQATVCARAQAQPPSPDQMWKCLIRLKTPLHGTPTLLYRLGAVQNEITITHPVTKY 1644

QY 683 IMTCNSADLEWVTS-----ACSGKPAIIPREVLYRE 714  
DB 1645 IMTCNSADLEWVTSYVWLVGGVLAALAAAYCLSTGCVVIVGRIVLSGKPAIIPREVLYQE 1704

QY 715 FDEMECSQHLPIYEQGMMLAEQFKOKALGL-----SRGKPAIYVDPKEVLYQQYDEMEEC 770  
DB 1705 FDEMECSQHLPIYEQGMMLAEQFKOKALGLLQTSRQAE-VITPAVQTWQKL-EVFWA 1762

QY 771 SQAPYIEQAQVIAHQFKEKVLGLINDQVV-----VTPDKELIYE----- 811  
DB 1763 KHMWNFISGQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLTTGTQLLFNLLGGWV 1816

QY 812 -----AFDEMECASKAALIEEGQRMALMSKIQGLLG----- 845  
DB 1817 AAQLAAPGAATAFVAGLAGAAGISVGLGKVLVDILAGYAGVAGALVAFKIMSGEVPST 1876

QY 846 -----ILRRHVPGEGAVQWMNRLIAFASRGNHVSPTHYVP 881  
DB 1877 EDLVNLLPALSPCALVGVVCAALLRRHVPGEGAVQWMNRLIAFASRGNHVSPTHYVP 1936

QY 882 S----- 882  
DB 1937 ESDAARVTAI LSSLVTQLRLRLHQLWISSECTTPCSGSMRLDIWDWICEVLSDFKTWLK 1996

QY 883 ----- 882  
DB 1997 AKLMPQLPGIPFVSCQRYGVWRGDGIMHTRCHGAEITGHVKNGTMRIVGPRTCRNMW 2056

QY 883 ----- 882  
DB 2057 SGTFFPINAYTGTCTPLPAPNYKFAIWRVSAEVEYTRRVGDFHYVSGMTNDLKCPCQI 2116

QY 883 -----RSRRFA-----QALPVWARPDPY----- 899

DB 2117 PSPEFFTELDGVRHLRHFADPKPCLLRREEVSRVGLHEYPVGSQPCPEPDAVLTSMLT 2176  
QY 900 -----NPP-----LVET---WKK- 909

DB 2177 DPSHITAEAGRRRLARGSPSSMASSASOLSAPSLKATCTANHDSPEALIEANLLWRQE 2236  
QY 910 -----PDYEPVPHG-----RSSRRFAQALFVWARPDPYNPP 940

DB 2237 MGNITRIVSESNKVVILDSFDPLVAEEDEREVSVAEILKRSRRFARALPVWARPDPYNPP 2296  
QY 941 LVETWKKPDYEPVPHVGRKTKNTNRPPQDVFPFGGQIVGRGPPPIPKARR 992  
DB 2297 LVETWKKPDYEPVPH-----GCPLPPRSPRPVPPPRK 2329

RESULT 5  
O36608 PRELIMINARY; PRT; 3011 AA.  
AC O36608;  
AD O36608;  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus strain H77.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.  
OX NCBI\_TaxID=63746;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H77;  
RX MEDLINE=972385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus  
RT are infectious when directly transfected into the liver of a  
RT chimpanzee."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).  
DR EMBL; AF011751; AAB67036.1; --  
DR PIR; A44150; A44150.  
DR PIR; PQ0804; PQ0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P27958; 1HEI.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_NS5b.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.



DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RDRP.  
DR InterPro; IPR001650; Helicase C.  
DR InterPro; IPR004109; Peptidase 329.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U35\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase C; 1.  
DR Pfam; PF00998; Viral\_RDRP; 1.  
DR SMART; SM00487; DEXDG; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327222 MW; 293F91430A0D4067 CRC64;  
  
Query Match 48.5%; Score 2868; DB 2; Length 3011;  
Best Local Similarity 51.74; Pred. No. 3.4e-163;  
Matches 616; Conservative 41; Mismatches 105; Indels 430; Gaps 17;  
  
QY 204 TGAARATTSGLTSLSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNDSPVPQSFQ 262  
DB 1165 SGGPLLCAGHAVGLFAAAYCTGRVAKAVDFIPVENLGTTRSPVFTDNDSPVPQSFQ 1224  
  
QY 263 VAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLGFAYMSKAHGIDPNIRTGVRT 322  
DB 1225 VAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLGFAYMSKAHGIDPNIRTGVRT 1284  
  
QY 323 ITTGSPITYSTYCKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDOAETAGARLV 382  
DB 1285 ITTGSPITYSTYCKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDOAETAGARLV 1344  
  
QY 383 VLATATPPGCVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 442  
DB 1345 VLATATPPGCVTVSHNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 1404  
  
QY 443 AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGVTGDFSDVIDCNTCTQTQVDF 502  
DB 1405 AKLVALGINAVAYRGLDVSVIPTSGDVVVVSTDALMTGVTGDFSDVIDCNTCTQTQVDF 1464  
  
QY 503 SLDPFTTETITLPQDAVSRTQRRGTRGKPGIYRFVAPGCRPSGMFDSVLCCEYDAG 562  
DB 1465 SLDPFTTETITLPQDAVSRTQRRGTRGKPGIYRFVAPGCRPSGMFDSVLCCEYDAG 1524  
  
QY 563 CAYELTPAETTVRLRAYMTNPGLPVQDHLFEWEGVFTGLTHIDAHFLSQTKQSGENLP 622  
DB 1525 CAYELTPAETTVRLRAYMTNPGLPVQDHLFEWEGVFTGLTHIDAHFLSQTKQSGENFP 1584  
  
QY 623 YLVAYQATVCARAQAPPPSDQWKKLIRLKTPLHGTPTLLYRLGAVQNEITITHTPVTKY 682  
DB 1585 YLVAYQATVCARAQAPPPSDQWKKLIRLKTPLHGTPTLLYRLGAVQNEITITHTPVTKY 1644  
  
QY 683 IMTCMSADLEVVTS-----ACSGKPAIIPDREVLRYE 714  
DB 1645 IMTCMSADLEVVTSVWLVGVLAAALAAAYCLSTGCVVIVGKRVLSGKPAIIPDREVLRYE 1704  
  
QY 715 FDEMECSQHLPIEQGMMLAEQPKQKALGL-----SRGGRPAIVPDKEVLYLQYDENEEC 770  
DB 1705 FDEMECSQHLPIEQGMMLAEQPKQKALGLLOATSRHAE-VITPAVQTNWQKL-EVFWA 1762

QY 771 SQAAPYIEQAQVIAHQFKEKVLGIDNDQV-----VTPDKELLYE----- 811  
DB 1763 KHMWNFISGIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLTTGOTLLFNILGGWV 1816  
  
QY 812 -----APDEMEECASKAALIEGQRMVEMLSKSIQGLLG----- 845  
DB 1817 AAQLAAPGAATAFVGAGLAGAAIGSVGLGKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876  
  
QY 846 -----ILRRHVGPGEAGAVQMMNRLIAFASRGNHVSPTHYVP 881  
DB 1877 EDLVNLLPALSPGALVGVVCAAILRRHVGPGEAGAVQMMNRLIAFASRGNHVSPTHYVP 1936  
  
QY 882 S----- 882  
DB 1937 ESDAAARVTAISSLTVTQLLRHLQHWISSECTTTCGSGSWLRDWDWICEVLSDFKTWLK 1996  
  
QY 883 ----- 882  
DB 1997 AKLMPOLPGTFVSCQGYRGVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPRTCRNMW 2056  
  
QY 883 ----- 882  
DB 2057 SGTFPINAVTTGPCTPLPAPNYKFAWVSAEYVEIRRVGDHYVSGMTTDLNLCPCQI 2116  
  
QY 883 -----RSRRFA-----QALPVWARPDY----- 899  
DB 2117 PSPEFTTELDGVRHLRFAPCKPELLREEVSVFRVGLHEYPVGSQLPCEPEPDVAULTSMLT 2176  
  
QY 900 -----NPP-----LVET---WKK- 909  
DB 2177 DPSHITAEAGRRRLARGSPSPMASSASQLSAPSLKATCATANHDSPAELIEANLLWROE 2236  
  
QY 910 -----PDYEPVHVH-----RSSRRFAQALPVWARPDPVNP 940  
DB 2237 MGNITRVSESNKVVILDSFDPLVAEDEREVSVPAILKSRFRARALPVWARPDPVNP 2296  
  
QY 941 LVETWKPKDYPPEPVHVRGKTKNTNRPPQDVKPPGGQIVGRRGPPPIPKARR 992  
DB 2297 LVETWKPKDYPPEPVH-----GCPLPPPPSPVPPPRK 2329  
  
RESULT 7  
Q9PMW9 PRELIMINARY; PRT; 3015 AA.  
ID Q9PMW9  
AC Q9PMW9  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99420396; PubMed-10489358; DOI=10.1006/viro.1999.9889;  
RA Yanagi M, Purcell R.H., Emerson S.U., Bukh J.;  
RT "Hepatitis C virus; an infectious molecular clone of a second major  
RT genotype (2a) and lack of viability of intertypic 1a and 2a  
RT chimeras.";  
RL Virology 262:250-263(1999).  
DR EMBL; AF177039; AA01181.1; -;  
DR EMBL; AF177037; AA01179.1; -;  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P27958; 1HEI.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.





```
AC Q6IX04;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-HCV1a.
RA Brann T.W., Kottilil S., Polis M., Imamichi T.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY615798; AAT44836.1; -.
DR HSSP; P27958; 1A1V.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019031; C: viral capsid; IEA.
DR GO: GO:0019038; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO: GO:0016787; F: hydrolase activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0006350; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01003; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; HXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.
FT NON_TER 2908
SQ SEQUENCE 2908 AA; 315737 MW; BF5A4BC591498A4F CRC64;

Query Match 48.5%; Score 2866; DB 2; Length 2908;
Best Local Similarity 52.3%; Pred. No. 4.2e-163;
Matches 610; Conservative 38; Mismatches 99; Indels 419; Gaps 17;

QY 224 SQNIQLIVDFIVENLETTMRSPPVTPQSFQVAHLHAPTSGSKTKVPAAYA 283
DB TRGVAKADVFIVENLETTMRSPPVTPDSSPPVTPQSFQVAHLHAPTSGSKTKVPAAYA 1245
QY 284 AQGYKVLVLPNSVAATLGFAYMSKAHGIDPNRTGVRTITGSPITYSYGKFLADGGC 343
```

```
Db 1246 AQGYKVLVLPNSVAATLGFAYMSKAHGIDPNRTGVRTITGSPITYSYGKFLADGGC 1305
QY 344 SGGAYDIIICDECHSTDATSIIGITGTVLDOAETAGARLVVLATATPPGCVTPHPNIEEV 403
Db 1306 SGGAYDIIICDECHSTDATSIIGITGTVLDOAETAGARLVVLATATPPGCVTPHPNIEEV 1365
QY 404 ALSTTGIPYGVKAIPLVFKGGRHLIFCHSKKCKDELAALKVALGINAVAYVGLDVSV 463
Db 1366 ALSTTGIPYGVKAIPLVFKGGRHLIFCHSKKCKDELAALKVALGINAVAYVGLDVSV 1425
QY 464 IPTSGDVVVVATDALMTGYTGDVDFSDVDCNTCTVTQTVDFSLDPTFTTETITLPODAVSRT 523
Db 1426 IPTSGDVVVVATDALMTGYTGDVDFSDVDCNTCTVTQTVDFSLDPTFTTETITLPODAVSRT 1485
QY 524 QRRGRTGRGPGIYRFVAPGERPSGMFSSVLCSCYDAGCAWYELTTPAETTVRLRAYNNT 583
Db 1486 QRRGRTGRGPGIYRFVAPGERPSGMFSSVLCSCYDAGCAWYELTTPAETTVRLRAYNNT 1545
QY 584 PGLPVCQDHLEFWEQVFTGLTHDAHLSQTKSGENLPYLVAQATVCABAQAPPPSMD 643
Db 1546 PGLPVCQDHLEFWEQVFTGLTHDAHLSQTKSGENLPYLVAQATVCABAQAPPPSMD 1605
QY 644 QMWKCLIRLKPTRLHGPTLLYRLGAVONEITLTHPVTKYIMTCSADLEVVTS----- 696
Db 1606 QMWKCLIRLKPTRLHGPTLLYRLGAVONEITLTHPVTKYIMTCSADLEVVTS----- 1665
QY 697 -----ACSGKPAIIPDREVLYREFDEMECECSOHLPIYIQGMMLA 735
Db 1666 VLAALAAAYCLSTGCVGIIGRIVLSGKPAVIPDREVLYREFDEMECECSOHLPIYIQGMMLA 1725
QY 736 EQFKQKALGL-----SRGKPAIPDKEVLQOYD-----EMEECSQAAPYIEQAQVIAHQ 786
Db 1726 EQFKQKALGLLQATASQAE-VIAPAVQTNMQKLEAFWAKHMMNFISGIYLAGLSTLPG- 1783
QY 787 FKEKVLGLIDNDQVVVTP---DKEILYE-----AFDEMECEASKAALI 826
Db 1784 -NPAIASLMAPTAATAVTSPLTTSQTLFLNLGSWAAQLAAPCAATAFVGAGLAGNAIGSV 1842
QY 827 BEGQMAEMLKSKIQGLLG-----IL 847
Db 1843 GLGKVLVDILAGYGAGVAGALVAFKINSGEPPPTEDLVNLLPALILSPGALVGVVCAAIL 1902
QY 848 RRHVGPGEQAVQWNRLLIAFASRGNHVSPTHYVPS----- 882
Db 1903 RRHVGPGEQAVQWNRLLIAFASRGNHVSPTHYVPSDAAARVTAISSLTVTXLLRLHQ 1962
QY 883 ----- 882
Db 1963 WVSSECTTPCSGSLRDIWDWICEVLSDFKTLWAKLMPQLPGIPFVSCQRYGVRWGRD 2022
QY 883 ----- 882
Db 2023 GIMHTRCHGAEITGHVKNGTMRIVGPKTCRNWMSGTTPINAYTTGPTCTPLPAPNYTAL 2082
QY 883 -----RSRRA----- 888
Db 2083 WRVSAEYVEIRRVGDFHYVTGMTADNLKPCQVPSPEFFTELDGVRHLRFAPCKPLLR 2142
QY 889 -----QALPVWARPDY-----NPP----- 902
Db 2143 DEVSFRVGLHDYPVGSQLPCEPEPDVAVLTSMLTDPSSHITAAARRRLARGSPPSSEASS 2202
QY 903 -----LVET---WKK-----PDYEPVPHV 918
Db 2203 ASQLSAPSLRATCTTNHDSPDALIEANLLWRQEMGNITRVESKNVILDSFPLVAE 2262
QY 919 G-----RSSRRFAQLPVWARPDPYVNPPLVETWKKPDYEPVPHVHGKTKRNTNR 966
Db 2263 EDEREVSVAABILKRRRFTPALPIWARPDPYVNPPLLEEWKKPDYEPVPHVHGKTKRNTNR 2317
QY 967 RPQDVKFGGQIVGRGPPPIPKARR 992
```



Db 2318 PPQ-----SPPVPPPRK 2329

RESULT 10

O36609 PRELIMINARY; PRT; 3011 AA.

AC O36609; (T-EMBLrel. 05, Created)

DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)

DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

DE Polyprotein.

OS Hepatitis C virus strain H77.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.

OX NCBI\_TaxID=63746;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=H77;

RX MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;

RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;

RT "Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee.";

RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).

RL ENBL; AF011752; AAB67037.1; -.

DR PIR; A44150; A44150.

DR PIR; P00804; P00804.

DR PIR; P03026; P03026.

DR PIR; P03027; P03027.

DR PIR; P03028; P03028.

DR HSP; P26664; 1HEI.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; P:ATP binding; IEA.

DR GO; GO:0008026; P:ATP-dependent helicase activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO; GO:0019087; P:viral transformation; IEA.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD/DEAH\_N.

DR InterPro; IPR011545; DEAD/DEAH\_N.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV env.

DR InterPro; IPR002519; HCV NS1.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRp.

DR InterPro; IPR001650; Helicase C.

DR InterPro; IPR004109; Helicase C.

DR InterPro; IPR009003; Pept\_Ser\_Cys.

DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR SMART; SM00487; DEXDg; 1.

DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.

SQ SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA9B5DC CRC64;

Query Match 48.4%; Score 2861; DB 2; Length 3011;  
Best Local Similarity 51.6%; Pred. No. 8.9e-163;  
Matches 615; Conservative 40; Mismatches 107; Indels 430; Gaps 17;

QY 204 TGGAAATTSGLTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNSSPPVQSFQ 262  
DB 1165 SGGPLLCPRAGHAVGLFRAAVCTRGVAKAVDFIPVENLGTTRSPVFTDNSSPPVQSFQ 1224

QY 263 VAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVRT 322  
DB 1225 VAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVRT 1284

QY 323 ITTGSPIITYGKFLADGCGSGGAYDIIICDECHSTDATSIILGIGTVLDQATAGARLV 382  
DB 1285 ITTGSPIITYGKFLADGCGSGGAYDIIICDECHSTDATSIILGIGTVLDQATAGARLV 1344

QY 383 VLATATPPGSGVTPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKDELA 442  
DB 1345 VLATATPPGSGVTPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKDELA 1404

QY 443 AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVTQTVDF 502  
DB 1405 AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVTQTVDF 1464

QY 503 SLDPFTTITITLPODAVSRQRRGTGRGKPGIYRFVAPGERPSGFMFSSVLCYCYDAG 562  
DB 1465 SLDPFTTITITLPODAVSRQRRGTGRGKPGIYRFVAPGERPSGFMFSSVLCYCYDAG 1524

QY 563 CAMELTTPAETTVRLRAYMNTGCLPVCQDHLFEWEGVFTGLTHIDAHFLSOTKQSGENLP 622  
DB 1525 CAMELTTPAETTVRLRAYMNTGCLPVCQDHLFEWEGVFTGLTHIDAHFLSOTKQSGENFP 1584

QY 623 YLVAYQATVCARAQAPPSQDMQMKCLIRLKPETHGTPPLLRLGAVQNEITLTHPVTKY 682  
DB 1585 YLVAYQATVCARAQAPPSQDMQMKCLIRLKPETHGTPPLLRLGAVQNEITLTHPVTKY 1644

QY 683 IMTMSADLEVTST-----ACSGKPAIIPREVLYRE 714  
DB 1645 IMTMSADLEVTSTVWLGVGVAALAAAYCLSTGCVVIVGRIVLSGKPAIIPREVLYQE 1704

QY 715 FDEMEECQHLPIYIEQGMMAEQFKOKALGL-----SRGKPAIVPDKVELVYQYDEMEEC 770  
DB 1705 FDEMEECQHLPIYIEQGMMAEQFKOKALGLQOTASRAHE-VITPAVQTNQKL-EVFWA 1762

QY 771 SQAAPYIEQAQVIAHQFKEKVLGLDNDQV-----VTPDKELIYE----- 811  
DB 1763 KHMWNFISGIQVLG-----LSTLPNPAIASLMAFTAATAVTSPLTTGOTLLFNILGGWV 1816

QY 812 -----AFDEMEECASKAALIEEQGVMAEMLSKSIQGLLG----- 845  
DB 1817 AAQLAAPGAATAFVGAGLAGAAIGSVGLGKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876

QY 846 -----ILRRHVGPGEQAVQVMNRLIAFASRGNHVSPTHYVP 881  
DB 1877 EDLVNLLPAILSPGALVGVVCAAILRRHVGPGEQAVQVMNRLIAFASRGNHVSPTHYVP 1936

QY 882 S----- 882  
DB 1937 ESDVAARVTAISSLTVTLRLRLHQWISSECTTPCGSGSWLRDWDWICEVLSDFKTLWK 1996

QY 883 ----- 882  
DB 1997 AKLMPQLPGIPFVSCQRYGVWRGDGIMHTRCHCGAEITHGVKNGAMRIVGPRTCRNMW 2056

QY 883 ----- 882  
DB 2057 SGTFLINAYTGPCTPLPAPNYKALWRVSAEYVEIRRVGDFHYISGMTDNLKCPQCI 2116

QY 883 -----RSRRFA-----QALPVWARPDY----- 899



```
Db 2117 PSPEFFTELDGRLRRFPAPCKPLLREEVSRVGLHEYPVGSQLPCEPEPDVAVLTSMLT 2176
Qy 900 -----NPP-----LVET---WKK- 909
Db 2177 DPSHITAEAGRLRAGSPSMASSSASQSLKATCTANHDSPDAELIEANLLNRQE 2236
Qy 910 -----PDYEPVVHG-----RSSRFAQALPVWARPDPNPP 940
Db 2237 MGGNITRVESKENVILDSFDPLVAEDEREVSVPAILRKSRFAPALPVWARPDPNPP 2296
Qy 941 LVETWKPDYEPVVHGRKTKVNRNRRPDQVKFGGQGVICVRCRPPPKARR 992
Db 2297 LVETWKPDYEPVVH-----GCPLPPRSPVPVPPPKR 2329

RESULT 11
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP Desai S.M., Devare S., Yamaguchi J.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF290978; AAG02099.1; --
DR PIR; A44150; A44150.
DR PIR; Q0804; Q0804.
DR PIR; P50326; P50326.
DR PIR; P50327; P50327.
DR PIR; P50328; P50328.
DR HSSP; P26664; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
```

```
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327108 MW; A6BECF5A3B3EE13F CRC64;

Query Match 48.2%; Score 2851; DB 2; Length 3011;
Best Local Similarity 51.5%; Pred. No. 3.5e-162;
Matches 611; Conservative 43; Mismatches 113; Indels 420; Gaps 17;

Qy 204 TGGAAARTTSLGTSLSFPG-ASQNIQLIVDFIPVENLETTMRSVPVFTDNNSPVPVPSFQ 262
Db 1165 SGGPLLCAGHAVGLFRAAVCTRGAKAVDFIPVENLETTMRSVPVFTDNNSPVPVPSFQ 1224
Qy 263 VAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGDIPNIRTGVRT 322
Db 1225 VAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGVDPNIRTGVRT 1284
Qy 323 ITTGSPTITYTGKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDQAEATAGARLV 382
Db 1285 ITTGSPTITYTGKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDQAEATAGARLV 1344
Qy 383 VLATATPGSVTVPHPNIEEVALSTTGEIIPYKAIPLEVIKGRHLIFCHSKKCDELA 442
Db 1345 VLATATPGSVTVPHPNIEEVALSTTGEIIPYKAIPLEVIKGRHLIFCHSKKCDELA 1404
Qy 443 AKLVALGINAVAYYRGDVSIVPTSGDVVVVATDALMTGYTGDFDSVDCNTCVTQTVDF 502
Db 1405 AKLVALGINAVAYYRGDVSIVPASGDVVVSTDALMTGTFDPPVIDCNTCVTQTVDF 1464
Qy 503 SLDPTFTIETITLPQDAVSRVTRGRGKPGIYRFVAPGERPSGMPDSSVCECYDAG 562
Db 1465 SLDPTFTIETITLPQDAVSRVTRGRGKPGIYRFVAPGERPSGMPDSSVCECYDAG 1524
Qy 563 CAWYELTPAETTVRLRAYMNTPLGVCDHLEFEGVFTGLTHDAHFLSOTKSGENLP 622
Db 1525 CAWYELTPAETTVRLRAYMNTPLGVCDHLEFEGVFTGLTHDAHFLSOTKSGENFP 1584
Qy 623 YLVAYQATVCARAQAPPPSDQMKLIRLXPTLHGPTPLLYRLGAVQNEITLTHPVTKY 682
Db 1585 YLVAYQATVCARAQAPPPSDQMKLIRLXPTLHGPTPLLYRLGAVQNEITLTHPVTKY 1644
Qy 683 IMTCMSADLEVVTS-----ACSGKPAIIPDREVLVYRE 714
Db 1645 IMTCMSANPEVTVSTWVLVGGVLAALAAAYCLSTGCVIVGRIVLSGKPAIIPDREVLVYRE 1704
Qy 715 FDEMECSQHLPIYIEQGNMLAEQFKQKALGL-----SRGKPAIIPDKEVLVYQYD----- 765
Db 1705 FDEMECSQHLPIYIEQGNMLAEQFKQKALGL-----SRGKPAIIPDKEVLVYQYD----- 1763
Qy 766 EMECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVPVTP---DKEILYE----- 811
Db 1764 HMMVFISGTQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTLTFLNLLGGWVAQA 1821
Qy 812 -----AFDEMECASKAALIEEGORMAEMLKSKIQLGL----- 845
Db 1822 APGAATAPVAGLAGAAGATGVSGLGKVLVDILAGYAGVAGVALVAFKIMSGEVPSTEDLVN 1881
Qy 846 -----ILRRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPS----- 882
Db 1882 LLPAILSPGALVGVVCAAILRRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPSDA 1941
Qy 883 ----- 882
Db 1942 ARVTAIISLNTVTLRLRLHQLHGWISCTTPCSGSLRDIWDWICEVLSDFKTLKAKLMP 2001
```

```
Qy 883 ----- 882
Db 2002 QLPQIPFVSCQGVYGVWVGIMHTRCHCAEITGHVKNGTMRIVGPRTCRNWWSGTFP 2061
Qy 883 ----- 882
Db 2062 INAYTGPCTPLPAPNYKFALMRVSAEYVEIRRVGDFHYVSGWTTDNLKCPCQIPSPF 2121
Qy 883 ----- 882
Db 2122 FTELDGVLRHFAFPCKPLLRVEVSFRVGLHYPVGSQLPCEPEPDVAVLTSMLTDPHSI 2181
Qy 900 ----- 909
Db 2182 TABAAGRRLARGSPPPSMASASSASOLSAPLSKATCTTNHSDPAELIEANLLWRQWGGNI 2241
Qy 910 ----- 945
Db 2242 TRVESENKVILDSFDPLVAEEDEREVSVAELRKSORFARALFWARPDYNPLLIETW 2301
Qy 946 KKPDYEPVHVHGRKTRNRRPDQVKFPGGGQIVGRRGPPIPKARR 992
Db 2302 KEDYEPVHVH-----GCLPPRSPVPPPRK 2329

RESULT 12
Q03463
ID Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463
DT 01-NOV-1996 (TremBurel. 01, Created)
DT 01-NOV-1996 (TremBurel. 01, Last sequence update)
DT 01-MAR-2004 (TremBurel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishihiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H., Teada F., Miyakawa Y., Mayumi M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177 (1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchauspe G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255 (1994).
RN [5]
RP SEQUENCE FROM N.A.
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; D10749; BAA01582.1; -.
```

```
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR PIR; S40770; S40770.
DR HSSP; P26664; IHEI.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_Core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Peptidase_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327114 MW; 97E9052C0250463B CRC64;

Query Match 48.2%; Score 2850; DB 2; Length 3011;
Best Local Similarity 51.2%; Pred. No. 4.1e-162;
Matches 608; Conservative 44; Mismatches 115; Indels 420; Gaps 18;

Qy 204 TGGAAARTTSGTLTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDSSPPVPSQSFQ 262
Db 1165 SGGLPLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLETTMRSPVFTDSSPPVPSQSFQ 1224
Qy 263 VAHLHAPTGSKSTKVPAAAYAAQGYKVLVNLNPSVAATLFGAYMYSKAHGIDPNIRTVRT 322
Db 1225 VAHLHAPTGSKSTKVPAAAYAAQGYKVLVNLNPSVAATLFGAYMYSKAHGIDPNIRTVRT 1284
Qy 323 ITTGSPITTYGKFLADGCGSGGAYDIIICDCHSTDATSLIGITVLDQAEATGARLV 382
Db 1285 ITTGSPITTYGKFLADGCGSGGAYDIIICDCHSTDATSLIGITVLDQAEATGARLV 1344
Qy 383 VLATATPPGSIPTVPHNIEVALSTTGEIPFYKKAIPLEVIKGRHLIFCHSKKKCDELA 442
Db 1345 VLATATPPGSIPTVPHNIEVALSTTGEIPFYKKAIPLEVIKGRHLIFCHSKKKCDELA 1404
Qy 443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVIDCNTCVTQTVDF 502
```

Db	1405	AKLVAGNAYVYRGLDVSIPYSGDGVVVVATDALMTGYTGDFSDSIDCNCVTQTQVDF	1464
Qy	503	SLDPTFTIETITLQDQAVSRQGRGRGPGIYRFVAPGRPSGMFSDSVLCECYDAG	562
Db	1465	SLDPTFTIETITLQDQAVSRQGRGRGPGIYRFVAPGRPSGMFSDSVLCECYDTG	1524
Qy	563	CWYELTPAETVRLRAYMNTPGIPVQCQDHLFEFWEVFTGLTHIDAHFLSQTQSGENLP	622
Db	1525	CWYELTPAETVRLRAYMNTPGIPVQCQDHLFEFWEVFTGLTHIDAHFLSQTQSGENFP	1584
Qy	623	YLVAQATVCARAQAPPSDWMKCLIRLKPETHGPTLLYRLGAVQNEITLHPVTKY	682
Db	1585	YLVAQATVCARAQAPPSDWMKCLIRLKPETHGPTLLYRLGAVQNEITLHPVTKY	1644
Qy	683	IMTWSADLEVVTTS-----ACSGKPAIIPDREVLYRE	714
Db	1645	IMTWSADLEVVTTSVWLVGGVLAALAYCLSTGCVVIVGRIVLSGRPAIIPDREVLYRE	1704
Qy	715	FDEMECSQHLPIYIEQGMMMLAEQFKQKALGL-----SRGGKPAIIPDKEVLYQQYD-----	765
Db	1705	FDEMECSQHLPIYIEQGMMMLAEQFKQKALGLLQTASRQAE-VIAPTQVQTNQKLEAFWAK	1763
Qy	766	EMECSQAPYIEQAQVTAHQFKKVLGLINDQVWVTP-----DKEILYE-----	811
Db	1764	HMNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTLNFIILGGWVAQA	1821
Qy	812	-----AFDEMECASKAALTEEGORMAEMLSKIQGLLG-----	845
Db	1822	APGAATAFVGLSLAGAAVGSVGLGRVLDIILAGYAGVAGALVAFKIMSGELPSTEDLVN	1881
Qy	846	-----ILRRHVGPGEQAVQWMMRLIAFASRGNHVSPTHYVPS-----	882
Db	1882	LLPALSPGALVGVVCAAILRRHVGPGEQAVQWMMRLIAFASRGNHVSPTHYVPSDAA	1941
Qy	883	-----	882
Db	1942	ARVTAISSLVTQTLLRRLHOMLSSESTPCSGSWLRDINDWICEVLSDFKTLKTLKMP	2001
Qy	883	-----	882
Db	2002	HLGIPFVSCQHYGKVGWRGDGIMHTRCHGAEITGHVKNGTMRIVGPKTCRNWMSGTFP	2061
Qy	883	-----	882
Db	2062	INAYTGPCTPLPAPNYTFALWRVSAEYVEIRRVGDFFHYVTGMTDNLKCPQVPSPEF	2121
Qy	883	-----RSRPA-----QALPVWARPDY-----	899
Db	2122	FTELDGVRLHRFAPPCKPLLEEVSVFRVGLHDYPVGSQLPCEPEFDVAVLTSMLTDP	2181
Qy	900	-----NPP-----LVET---WKK-----	909
Db	2182	TAAAGRLRAGSPSESSASSQSLSKATCTINHDSPDAELTEANLLWRQMGNGI	2241
Qy	910	-----PDYEPVPHVG-----RSSRRFAQALPVWARPDPYNPLVETW	945
Db	2242	TRVSENKVILDSFDPLVAEEDEREISVPAEILKSRRTQALPIMARPDPNPLIETW	2301
Qy	946	KKPDYEPVPHVGRKTRNRRNQDKFPFGGQIVGRRGPPIPKARR	992
Db	2302	KKPNYEPVPHVGCLP-----PPQ-----SPVPVPPRK	2329
RESULT 13			
Q9DIT6			
ID	Q9DIT6	PRELIMINARY;	PRT; 3011 AA.
AC	Q9DIT6;		
DT	01-MAR-2001	(TReMBLrel. 16, Created)	
DT	01-MAR-2001	(TReMBLrel. 16, Last sequence update)	
DE	01-MAR-2004	(TReMBLrel. 26, Last annotation update)	
OS	Genomic RNA for polyprotein gene.		
	Hepatitis C virus.		

OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX	NCBI_TaxID=111103;
KN	[1]
RP	SEQUENCE FROM N.A.
EX	MEDLINE=21014673; PubMed=11115058;
RA	Kumar U., Tuthill T., Thomas H.C., Monjardino J.,
RT	"Sequence, expression and reconstruction of an HCV genome from a
RT	British isolate derived from a single blood donation.";
RL	J. Viral Hepat. 7:459-465(2000).
DR	EMBL; AJ278830; CAC03609.1; ..
DR	PIR; PS0326; PS0326.
DR	PIR; PS0327; PS0327.
DR	PIR; PS0328; PS0328.
DR	HSSP; P27958; 1A1V.
DR	GO; GO:0016021; C:integral to membrane; IEA.
DR	GO; GO:0019028; C:viral capsid; IEA.
DR	GO; GO:0019031; C:viral envelope; IEA.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR	GO; GO:0003723; F:RNA binding; IEA.
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR	GO; GO:0005198; F:structural molecule activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	GO; GO:0006350; P:transcription; IEA.
DR	GO; GO:0019079; P:viral genome replication; IEA.
DR	GO; GO:0019087; P:viral transformation; IEA.
DR	InterPro; IPR000345; CytC_heme_BS.
DR	InterPro; IPR001410; DEAD.
DR	InterPro; IPR011545; DEAD/DEAH_N.
DR	InterPro; IPR011492; Flavi_DEAD.
DR	InterPro; IPR002522; HCV capsid.
DR	InterPro; IPR002521; HCV_core.
DR	InterPro; IPR002519; HCV_env.
DR	InterPro; IPR002531; HCV_NS1.
DR	InterPro; IPR000745; HCV_NS4a.
DR	InterPro; IPR001490; HCV_NS4b.
DR	InterPro; IPR002868; HCV_NS5a.
DR	InterPro; IPR002166; HCV_RdRP.
DR	InterPro; IPR001650; Helicase_C.
DR	InterPro; IPR004109; Peptidase_S29.
DR	InterPro; IPR009003; Pept_Ser_Cys.
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.
DR	InterPro; IPR007095; RNA_pol_DS_PS.
DR	InterPro; IPR007094; RNA_pol_PSVir.
DR	Pfam; PF07652; Flavi_DEAD; 1.
DR	Pfam; PF01543; HCV_capsid; 1.
DR	Pfam; PF01542; HCV_core; 1.
DR	Pfam; PF01539; HCV_env; 1.
DR	Pfam; PF01560; HCV_NS1; 1.
DR	Pfam; PF01538; HCV_NS2; 1.
DR	Pfam; PF02907; HCV_NS3; 1.
DR	Pfam; PF01006; HCV_NS4a; 1.
DR	Pfam; PF01001; HCV_NS4b; 1.
DR	Pfam; PF01506; HCV_NS5a; 1.
DR	Pfam; PF00271; Helicase_C; 1.
DR	Pfam; PF00998; Viral_RdRP; 1.
DR	SMART; SM00487; DEXDC; 1.
DR	PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; Transmembrane.
KW	CHAIN 810 1026
FT	CHAIN 1027 1657 non-structural protein 2.
FT	CHAIN 1658 1711 non-structural protein 3.
FT	CHAIN 1712 1972 non-structural protein 4a.
FT	CHAIN 1973 2420 core protein.
FT	CHAIN 2421 3011 non-structural protein 4b.
FT	CHAIN 192 383 non-structural protein 5a.
FT	CHAIN 384 809 non-structural protein 5b.
FT	CHAIN 3011 AA; 327406 MW; 7B6264A74A5452D3 CRC64;
SEQUENCE	3011 AA; 327406 MW; 7B6264A74A5452D3 CRC64;

Query Match		48.0%; Score 2835; DB 2; Length 3011;
Best Local Similarity		50.7%; Pred. No. 3.3e-161;
Matches 612; Conservative 46; Mismatches 130; Indels 420; Gaps 18;		
QY	183	RTTSFGVSLFAPCAKQETHVTGGAARTTSGTSLTSPG-ASQNIQLIVDTPVENLET 241
DB	1144	RRDGRGSLSPRI SYLKGSSGGPLLCAGHAGVIFRAACVTRGAKAVEFPVENLET 1203
QY	242	TMRSPTFTDNSSPPVQSQVAHLHAPTSGKSTKVPAAAYAAQGYKVLVLPNSVAATLG 301
DB	1204	TMRSPTFTDNSSAPPVQSQVAHLHAPTSGKSTKVPAAAYAAQGYKVLVLPNSVAATLG 1263
QY	302	FGAYMSKAHGDINRTKVTITGSPITVSTYTKFLADGCGSGGAYDIIICDECHSTDA 361
DB	1264	FGAYMSKAHGDINRTKVTITGSPITVSTYTKFLADGCGSGGAYDIIICDECHSTDA 1323
QY	362	TSILGIGTGLDQAEAGARLWLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLE 421
DB	1324	TSILGIGTGLDQAEAGARLWLTATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLE 1383
QY	422	VIKGGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 481
DB	1384	VIKGGRHLIFCHSKKCKDDLAALVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 1443
QY	482	YTGDFOSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRTOGRGTGRGKGIYRFVA 541
DB	1444	FTGDFOSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRTOGRGTGRGKGIYRFVA 1503
QY	542	PGERPMSGDSVLCYDAGCAWYELTPAETTVLRAYNMTGRLPVCODHLEFWGVT 601
DB	1504	PGERPMSGDSVLCYDAGCAWYELTPAETTVLRAYNMTGRLPVCODHLEFWGVT 1563
QY	602	GLTHIDAHFLSOTKQSGENLPYLVAQVATCARAQAPPPSWDQWKCLRKLKPTLHGPTP 661
DB	1564	GLTHIDAHFLSOTKQSGENLPYLVAQVATCARAQAPPPSWDQWKCLRKLKPTLHGPTP 1623
QY	662	LLYRLGAVQNEITLTHPTVKYIMTCSADLEVTSS----- 696
DB	1624	LLYRLGAVQNEITLTHPTVKYIMTCSADLEVTSSVWLVGVLAAALAAAYCLSTGCVVIV 1683
QY	697	---ACSGKPAIIPDREVLREFDEMESCQHLPIYEQGMWLAOFKQKALGL-----SRGG 749
DB	1684	GRIVLSGKPIIPDQEVLYREFDEMESCQHLPIYEPGMWLAOFKQKALALLQTSRQA 1743
QY	750	KPAIVDPKEVLYQOYD-----EMEECSQAAPYEQAOVIAHQFKEKVLGLINDQVVVTP 804
DB	1744	E-VIAVAVQTNWORLEAFWAKHWNFIISGQYLAGUSTLPG--NPALASLMATAVTS 1800
QY	805	---DKEILYE-----AFDEMECASKAALIEEGORMAEMLSKIQGLL 844
DB	1801	LTTSTQLLFNLIGWVAAQAAPGAATAFVGAGLAGAAIGSVGLGVLDILAGYGAGVA 1860
QY	845	G-----ILRRHYGPGEGAVQWNNRLI 865
DB	1861	GALVAFKIMSGEVPSTEDMVNLLPAILSPGALVGVVVCATILRRHYGPGEGAVQWNNRLI 1920
QY	866	AFASRGNHVSPTHVPS----- 882
DB	1921	AFASRGNHVSPTHVPSDAAARVAILSSLTVTQLRLHQWVSSSTTPCSGSWLRI 1980
QY	883	----- 882
DB	1981	WDWICEVLSDFTWLKAKLMPQLPGIPLVSCRGYRGVWGQGVMMTRCHCGAEITGHVK 2040
QY	883	----- 882
DB	2041	NGTMRIVGPKTCRNMMSGTFPINAYTTGCTPLPAPNYTFALMRVSAEYVEIRRVGDFH 2100
QY	883	-----RSRRFA-----QAL 891
DB	2101	YVTGMTDNLRCQVQPSPEFFTELGDVRLHRAFPCKPILLRDEVSFRVLGHDPVGSQ 2160
QY	892	PVWARPDY-----NPP----- 902
RESULT 14		
POLG_HCVH		
ID	POLG_HCVH	STANDARD; PRT; 3011 AA.
AC	P27958;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	25-OCT-2004 (Rel. 45, Last annotation update)	
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);	
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2	
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)	
DE	(EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)	
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein	
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein	
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].	
OS	Hepatitis C virus (isolate H) (HCV).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus.	
OX	NCBI_TaxID=11108;	
RN	[1]_TaxID=11108;	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=92052256; PubMed=1658800;	
RA	Inchauspe G., Zebede S., Lee D.H.H., Sugitani M., Nasoff M.,	
RA	Prince A.M.;	
RT	"Genomic structure of the human prototype strain H of hepatitis C	
RT	virus: comparison with American and Japanese isolates.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).	
RN	[2]	
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.	
RX	MEDLINE=97331322; PubMed=9187654;	
RA	Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;	
RT	"Structure of the hepatitis C virus RNA helicase domain.";	
RL	Nat. Struct. Biol. 4:463-467(1997).	
RN	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.	
RX	MEDLINE=98154321; PubMed=9493270; DOI=10.1016/S0969-2126(98)00010-0;	
RA	Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,	
RA	Murcko M.A., Lin C., Caron P.R.;	
RT	"Hepatitis C virus NS3 RNA helicase domain with a bound	
RT	oligonucleotide: the crystal structure provides insights into the mode	
RT	of unwinding.";	
RL	Structure 6:89-100(1998).	
CC	-1- FUNCTION: Protease NS2 is responsible for the cleavage of NS2-NS3.	
CC	-1- FUNCTION: Protease NS3 is responsible for the cleavage of NS3.	
CC	-1- NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B.	
CC	-1- FUNCTION: NS4A forms a complex with NS3 and is essential for the	
CC	activation of NS3.	
CC	-1- FUNCTION: NS5A seems to have a transcriptional activatory role.	
CC	-1- FUNCTION: NS5B is a RNA-dependent RNA polymerase that plays an	
CC	essential role in the virus replication.	
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral	
CC	precursor polyprotein. Hydrolysis of four peptide bonds in the P6	
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.	
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +	
CC	{RNA} (N).	
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a	
CC	lipoprotein envelope. The envelope consists of two proteins: E1	
CC	and E2. The nucleocapsid is a complex of protein C and mRNA.	

-1- PTM: The structural proteins C, E1 and E2 are produced by proteolytic processing by the host signal peptidases.

-1- SIMILARITY: Contains 1 peptidase S29 domain.

-1- SIMILARITY: Contains 1 peptidase U39 domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M67463; AAA45534.1; --

PIR; A36814; GNVVCH.

PDB; 1A18; X-ray; A/B=1017-1214, C/D=1676-1698.

PDB; 1A1V; X-ray; A=1192-1667.

PDB; 1HEI; X-ray; --

MEROPS; S29.001; --

DR MEROPS; U39.001; --

DR TRANSFAC; T04155; --

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR000745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRP.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR009003; Pept\_Ser\_Cys.

DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.

DR InterPro; IPR004109; PeptIdase\_S29.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.

DR SMART; SM00487; DEXDC; 1.

DR 3D-structure; ATP-binding; Coat protein; Core protein;

KW Envelope protein; Glycoprotein; Helicase; Hydrolase;

KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;

KW Serine protease; Transferrase; Transmembrane.

FT INIT\_MET 1 1

FT CHAIN 1 191

FT CHAIN 192 383

FT CHAIN 384 746

FT CHAIN 747 809

FT CHAIN 810 1026

FT CHAIN 1027 1657

FT CHAIN 1658 1711

FT CHAIN 1712 1972

FT CHAIN 1973 2420

FT CHAIN 2421 3011

FT TRANSMEM 347 369

FT ACT\_SITE 1083 1083

FT ACT\_SITE 1107 1107

FT ACT\_SITE 1165 1165

FT NP\_BIND 1230 1237

FT SITE 1316 1319

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

-----

-1- PTM: The structural proteins C, E1 and E2 are produced by proteolytic processing by the host signal peptidases.

-1- SIMILARITY: Contains 1 peptidase S29 domain.

-1- SIMILARITY: Contains 1 peptidase U39 domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M67463; AAA45534.1; --

PIR; A36814; GNVVCH.

PDB; 1A18; X-ray; A/B=1017-1214, C/D=1676-1698.

PDB; 1A1V; X-ray; A=1192-1667.

PDB; 1HEI; X-ray; --

MEROPS; S29.001; --

DR MEROPS; U39.001; --

DR TRANSFAC; T04155; --

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR000745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRP.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR009003; Pept\_Ser\_Cys.

DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.

DR InterPro; IPR004109; PeptIdase\_S29.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.

DR SMART; SM00487; DEXDC; 1.

DR 3D-structure; ATP-binding; Coat protein; Core protein;

KW Envelope protein; Glycoprotein; Helicase; Hydrolase;

KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;

KW Serine protease; Transferrase; Transmembrane.

FT INIT\_MET 1 1

FT CHAIN 1 191

FT CHAIN 192 383

FT CHAIN 384 746

FT CHAIN 747 809

FT CHAIN 810 1026

FT CHAIN 1027 1657

FT CHAIN 1658 1711

FT CHAIN 1712 1972

FT CHAIN 1973 2420

FT CHAIN 2421 3011

FT TRANSMEM 347 369

FT ACT\_SITE 1083 1083

FT ACT\_SITE 1107 1107

FT ACT\_SITE 1165 1165

FT NP\_BIND 1230 1237

FT SITE 1316 1319

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

FT	CARBOHYD	305	305	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	417	417	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	423	423	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	430	430	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	448	448	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	476	476	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	532	532	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	540	540	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	556	556	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	576	576	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	623	623	N-linked (GlcNAc. . .)	(Potential)
FT	CARBOHYD	645	645	N-linked (GlcNAc. . .)	(Potential)
FT	STRAND	1032	1035		
FT	HELIX	1039	1048		
FT	STRAND	1050	1050		
FT	STRAND	1059	1063		
FT	STRAND	1068	1074		
FT	TURN	1075	1076		
FT	STRAND	1077	1081		
FT	HELIX	1082	1085		
FT	TURN	1086	1087		
FT	STRAND	1090	1092		
FT	TURN	1093	1094		
FT	STRAND	1095	1097		
FT	STRAND	1101	1103		
FT	TURN	1104	1107		
FT	STRAND	1108	1112		
FT	TURN	1115	1116		
FT	STRAND	1120	1120		
FT	STRAND	1122	1122		
FT	TURN	1129	1133		
FT	TURN	1135	1136		
FT	STRAND	1139	1144		
FT	TURN	1147	1148		
FT	STRAND	1149	1157		
FT	HELIX	1158	1161		
FT	TURN	1162	1163		
FT	STRAND	1165	1166		
FT	TURN	1168	1170		
FT	STRAND	1172	1173		
FT	TURN	1176	1186		
FT	TURN	1187	1188		
FT	STRAND	1189	1197		
FT	HELIX	1198	1205		
FT	STRAND	1680	1687		
FT	STRAND	1693	1694		
SQ	SEQUENCE	3011	AA; 327142	MM; 772CBB29CCD94753	CRC64;

Query Match 47.7%; Score 2821; DB 1; Length 3011;

Best Local Similarity 51.3%; Pred. No. 2.3e-160;

Matches 611; Conservative 41; Mismatches 110; Indels 430; Gaps 17;

Qy	204	TGGAAATTSGTSLFSPG-ASQNIQLIVDFIPVENLETTMRS	PVFTDSSPPVQSFQ	262
Db	1165	SGGFLCPTGAVGLFRAAVCTRGVAKAVDFIPVENLETTMRS	PVFTDSSPPVQSFQ	1224
Qy	263	VAHLHPTGSGKSTKVPAAVAAQGYKVLNPSVAATLFGAYMS	KAHGIDPNRTGVRT	322
Db	1225	VAHLHPTGSGKSTKVPAAVAAQGYKVLNPSVAATLFGAYMS	KAHGVDPNRTGVRT	1284
Qy	323	ITTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATSIL	GIGTVLDAQETAGARLV	382
Db	1285	ITTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATSIL	GIGTVLDAQETAGARLV	1344
Qy	383	VLATATPGSVTVPHNIEEVALSTTGEIPYGAIPLEVIGKGR	HLIFCHSKKCCDELA	442
Db	1345	VLATATPGSVTVSHNIEEVALSTTGEIPYGAIPLEVIGKGR	HLIFCHSKKCCDELA	1404
Qy	443	AKLVALGINAVAYRGDLDSVPTSGDVVVVATDALMTGYT	CGDSDSDCNTCVTQTVDF	502
Db	1405	AKLVALGINAVAYRGDLDSVPTSGDVVVVATDALMTGYT	CGDSDSDCNTCVTQTVDF	1464
Qy	503	SLDPTFTIETITLPQDAVSRTRGRGRKPGYRFVAPGERPS	GMFDSVLCCEYDAG	562.

```
Db 1465 SLDPTFTIETTLPODAVSTQRGRGTGRGKGIYRFVAPGERPSGDFSSVLCECYDAG 1524
Qy 563 CANYELTTPARTVRLRAYMNTTGLPCVODHLEFWEGVFTGLTHIDAHFLSOTKSGENLP 622
Db 1525 CANYELTTPARTVRLRAYMNTTGLPCVQDHLGFWEGVFTGLTHIDAHFLSOTKSGENFP 1584
Qy 623 YLVAYQATVCARAQAPPPSDMMKCLIRLKTPLHGTPTLLYRLGAVQNEITLTHPVTKY 682
Db 1585 YLVAYQATVCARAQAPPPSDMMKCLIRLKTPLHGTPTLLYRLGAVQNEITLTHPVTKY 1644
Qy 683 IMTCMSADLEVTS-----ACSGKPAIIPREVLYRE 714
Db 1645 IMTCMSADLEVTSVTLVGGVLAALAAAYCLSTGCVVIVGRIVLSGKPAIIPREVLYQE 1704
Qy 715 FDEMESCQHLPIYEGMMMLAEQKOKALGL-----SRGKPAIVPDKVELYQYDSEMEEC 770
Db 1705 FDEMESCQHLPIYEGMMMLAEQKOKALGLLQATSRHAE-VITPAVQTNNQKL-EVFWA 1762
Qy 771 SQAAPYIEQAQVIAHOFKEKVLGLINDQVV-----VTPDKELIYE----- 811
Db 1763 KHMNFISGIQYLAG-----LSTLPCNPALIASIMAPTAAVTSPLTTGQTLLFNILGGMV 1816
Qy 812 -----ADMEBECASKALIEEGQMAEMLKSKIQGLLG----- 845
Db 1817 AAQLAAPGAATAFVGAGLAGAALDSVGLGKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876
Qy 846 -----ILRRHVGPEGAVQNMNRLIAFASRGNHVSPTHYVP 881
Db 1877 EDLVNLLPALSPGALAVGVFASILRRRVGPEGAVQNMNRLIAFASRGNHVSPTHYVP 1936
Qy 882 S----- 882
Db 1937 ESDAARVTAISSLVTQLRLRHQWISSECTPCSGSLRDINDWICEVLSDFKTWLK 1996
Qy 883 ----- 882
Db 1997 AKLMPQLPGIPFVSCQYRGVNRGDMHTRCHGAEIGHVKNGTMRIVGPRCTCKNMW 2056
Qy 883 ----- 882
Db 2057 SGTFFINAYTGTCTELPAPNYKFAWRVSAEYVEIRRVGDFHYVSGMTDNLKPCQI 2116
Qy 883 -----RSRFA-----QALPVWARPDI----- 899
Db 2117 PSPEFTELDGLVLRHFAFPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPDVAVLSMLT 2176
Qy 900 -----NPP-----LVET---WKK- 909
Db 2177 DPSHITAAGARRLARGSPSPMASSASQLSAPSLKATCATANHDSPPDAELIEANLLWRQE 2236
Qy 910 -----PDYEPVWHG-----RSSRFAQALPVWARPDIYNNP 940
Db 2237 MGNITRVSEKNKVIDSFDPLVAEEDEREVSVPAILKRSRRFAPALPVWARPDIYNPL 2296
Qy 941 LVETWKKPDYEPVWHGRTKRNTRRPQDVKPPGGQIVGRRGPPPIPKARR 992
Db 2297 LVETWKKPDYEPVWH-----GCPLPPSPPPVPPRK 2329

RESULT 15
Q9J3H0 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=MD22;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF207763; AAF65953.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0254; PQ0254.
DR PIR; PS0329; PS0329.
DR HSSP; Q8JYSL; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3010 AA; 32709 MW; 737BEF31E3C2B28D CRC64;

Query Match 47.1%; Score 2783.5; DB 2; Length 3010;
Best Local Similarity 48.3%; Pred. No. 4.1e-158;
Matches 590; Conservative 59; Mismatches 125; Indels 448; Gaps 14;

Qy 183 RTTSFGVSLFAPGAKQKQETHVTGGAARTTSGLTSLFSFG-ASQNTQLIVDFIPVENLET 241
Db 1144 RRGDGRGSLSPRPVSYLKGSSGGPLCLPLGHVVGIFRAAVCTRGVAKAVDFVPVSMET 1203
Qy 242 TMRSPVFTDNSSPPVPPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLG 301
Db 1204 TMRSPVFTDNSSPPVPPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLG 1263
Qy 302 FGAYMSKAGIDPNIRTNRTTITTSPTITSTYTGKFLADGCGSGGAYDIIICDECHSTDA 361
Db 1264 FGAYMSKAYGTFDPIRNTGRTITTTGAPITTYTGKFLADGCGSGGAYDIIICDECHSTDS 1323
```





1000